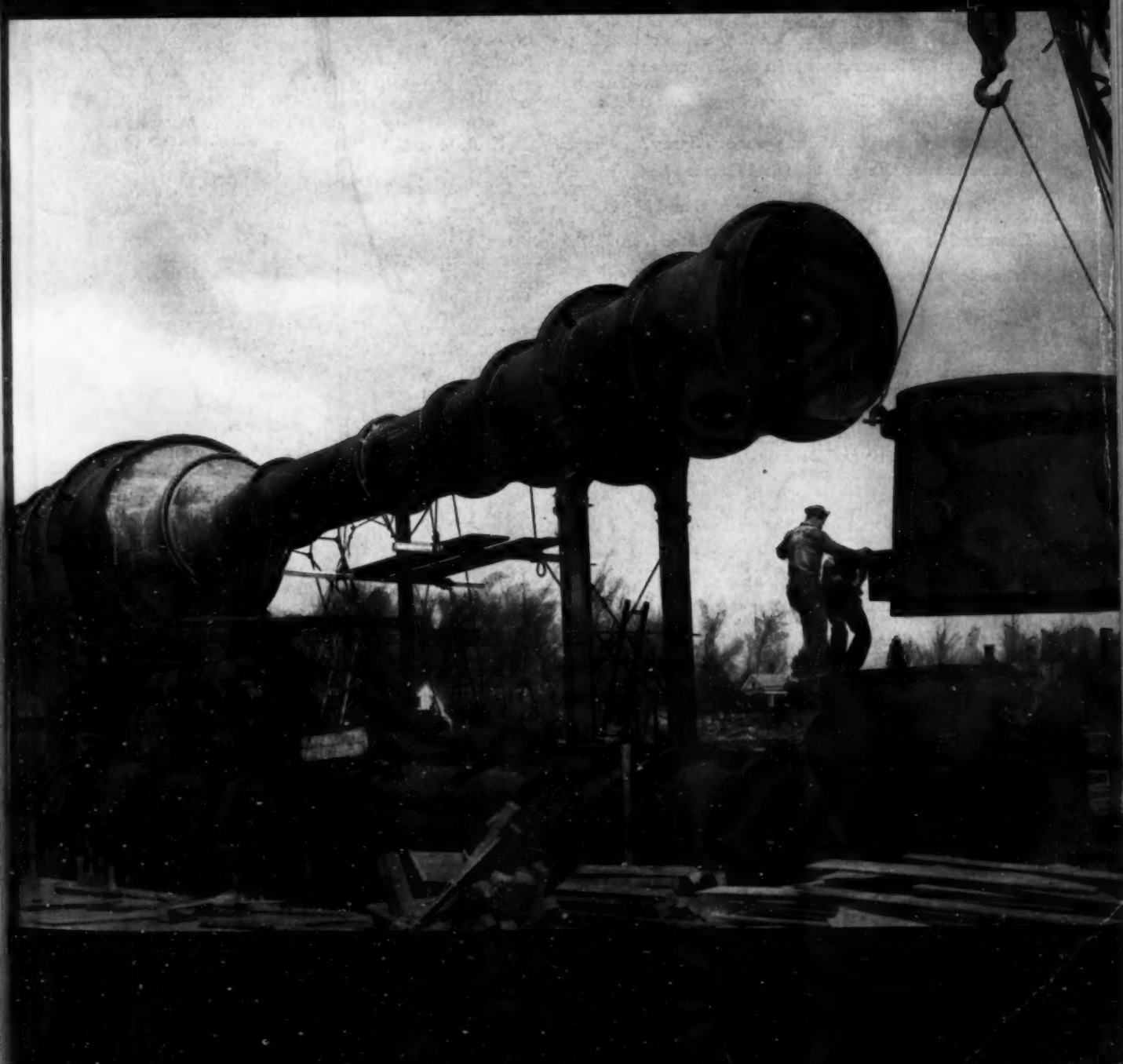
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 20, No. 10 OCTOBER, 1949

In This Issue

		Page
Ionization of Paraffin Wax by Beta-Radiation	ANDREW GEMANT	887
Antennas and Open-Wire Lines. Part II: Measurements on Two-Wire Lines	K. TOMIYASU	892
Effects of Stress and Deformation on the Martensite Transformation	ANDREW W. MCREYNOLDS	896
High Temperature Precision X-Ray Camera. Some Measurements of the Thermal Coefficients of Expansion of Beryllium	PAUL GORDON	908
Microwave *Q* Measurements in the Presence of Series Losses	L. MALTER AND G. R. BREWER	918
The Application of Non-Integral Legendre Functions to Potential Problems	R. N. HALL	925
Improved Spinning Top Homogeneous Spray Apparatus	K. R. MAY	932
Distribution of Temperature and Current in Cylindrical Bodies	CARL F. MUCKENHOUPT	939
Precipitation from Solid Solutions of C and N in α-Iron	CHARLES A. WERT	943
Theory of Growth of Spherical Precipitates from Solid Solution	CLARENCE ZENER	950
Infra-Red Spectra and Structure of Hevea and Gutta Elastomers	R. A. SAUNDERS AND D. C. SMITH	953
Magnetic Characteristics of an Oriented 50 Percent Nickel-Iron Alloy	J. H. CREDE AND J. P. MARTIN	966
Effect of Surface Roughness on Rolling Friction	J. J. BIKERMAN	971
Demodulation of a Frequency-Modulated Carrier and Random Noise by a Discriminator	NELSON M. BLACHMAN	976
Effective Use of Collimating Apertures in Small-Angle X-Ray Diffraction Cameras	ORVIL E. A. BOLDUAN AND RICHARD S. BEAR	983
Electron Microscope and Diffraction Study of Metal Crystal Textures by Means of Thin Sections	R. D. HEIDENREICH	993

Letters to the Editor:

Erratum: Determination of Relative Phases of Fourier Coefficients from X-Ray Diffraction Data	DAN MCLACHLAN, JR.	1011
Increase in *Q*-Value and Reduction of Aging of Quartz Crystal Blanks	A. C. PRICHARD, M. A. A. DRUESNE, AND D. G. MCCAA	1011
Addendum: Analysis of the Metal-Strip Delay Structure for Microwave Lenses	S. B. COHN	1011
Surface Polarization and Wettability of Inorganic Substances by Water	D. P. ENRIGHT, L. R. SONDERS, AND W. A. WEYL	1011
Index to Advertisers		xxxiv

BOARD OF EDITORS
ELMER HUTCHISSON, *Editor*

Associate Editors
R. BOWLING BARNES
OTTO BEECK
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
RAYMOND J. SEEGER
FREDERICK SEITZ
CHARLES S. SMITH
J. R. TOWNSEND

Publication of the
American Institute of Physics

OFFICERS
GEORGE R. HARRISON, *Chairman*
WALLACE WATERFALL, *Secretary*
GEORGE B. PEGRAM, *Treasurer*
JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD
LUIS W. ALVAREZ
J. W. BEAMS
WALLACE R. BRODE
E. U. CONDON
W. F. FAIR, JR.
HARVEY FLETCHER
GEORGE R. HARRISON
W. V. HOUSTON
PAUL KIRKPATRICK
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
J. R. OPPENHEIMER
GEORGE B. PEGRAM
HAROLD K. SCHILLING
JOHN T. TATE
WALLACE WATERFALL
M. W. WHITE
A. G. WORTHING

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
THEODORE VORBURGER, *Business Manager*
RUTH F. BRYANS, *Publication Manager*
BETTY HOFSTADTER, *Office Manager*
EILEEN B. NEUBERGER, *Circulation Manager*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $4 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices
Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 20, Number 10 October, 1949

Ionization of Paraffin Wax by Beta-Radiation

ANDREW GEMANT
Research Department, The Detroit Edison Company, Detroit, Michigan
(Received February 2, 1949)

This is a quantitative experimental study of the primary ionization of paraffin wax by beta-radiation. Electrons emitted by uranium and carbon 14 were used. Interpretation, by means of a simple theory, of experimental data on the electrical conductivity permits evaluation of the ionic mobility in paraffin and the efficiency of ion-pair production. Only small primary ions of relatively long mean life are detected by the conductivity method. Dipolar admixtures promote secondary reactions and lower the level of conductivities attained. No ionization was detected in two hydrocarbon polymers.

I

THIS paper presents the results of a study of the primary ionization of a hydrocarbon by beta-radiation. The ionization produced by beta-rays causes an increase in electrical conductivity, and it is this latter effect that was measured. Paraffin wax is particularly suited for the purpose because of its low inherent electrical conductivity, which does not mask the specific effect sought.[1]

The measuring cell is shown in Fig. 1. It was made of stainless steel, the sensitive electrode being insulated from the guard by an aniline-formaldehyde plastic. A thin layer of paraffin wax was formed at the bottom of the cell. The so-called high voltage electrode consisted of a thin layer of colloidal graphite (0.5 mg per cm^2). A metal ring placed on top of the graphite served as a connection to the d.c. source, a battery. The sensitive electrode was connected to an FP 54-electrometer circuit, the grid resistance being 1.6×10^{11} ohms.

The radioactive source material was spread on the bottom surface of a metal holder which fitted the metal ring as shown in the figure. The gap between the source and the graphite was about 0.5 mm. The area of the source was 20 cm^2, the same as that of the sensitive electrode.

Two sources were used in the experiments. One was uranyl nitrate. Although this material emits a combined alpha-, beta-, and gamma-radiation, it was ascertained by filter tests (to be described later) that ionization was due preponderantly to the 2.32-Mev electrons emitted by protactinium 234. To a good approximation, the same electrons cause most of the counts from a uranium sample as measured with a conventional thin-window Geiger counter.[2] By taking counts at different distances between the source and the Geiger tube, the number of 2.32-Mev betas emitted over a spatial angle of 2π was found to be 500 per sec. per cm^2. This figure will be used later in the evaluation of the results.

The second source was carbon 14 in the form of tridecanoic acid.[3] The active layer was about 0.5-mm thick; it emitted electrons (0.145 Mev) at a rate of 23×10^3 per sec. per cm^2 over an angle of 2π. Considering the absorption of these soft betas in the graphite, the number reaching the paraffin is reduced to 20×10^3 per sec. per cm^2. Because of the closeness of source and target, the geometry correction is small and was therefore ignored.

Three different thicknesses of paraffin were used, 0.015, 0.035, and 0.068 cm. Each layer was studied with both radioactive sources and at a number of different voltages.

For each experiment, the d.c. circuit was closed while the electron source was absent. The high charging current disappeared in about one hour; when the low residual current was reached, the radioactive source was placed upon the paraffin layer, the d.c. circuit remaining undisturbed. The current then gradually increased

[1] W. Scislowski, Acta Phys. Polonica **7**, 127 (1938).

[2] Andrew Gemant, J. App. Phys. **19**, 1160 (1948).

[3] Edward Hines and Andrew Gemant, Science **110**, 19 (1949).

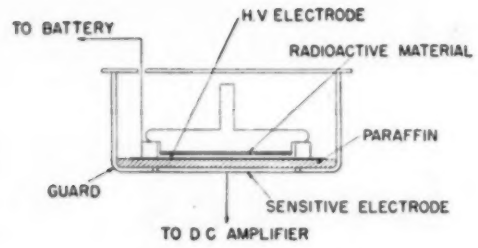


FIG. 1. Diagram of cell.

with time, approaching a constant value in about two hours. Next, the source was removed while the circuit was still closed. The current gradually decreased, approaching the original low value in another two hours.

II

The result, in terms of conductivities less the residual value, for one particular uranium run is shown

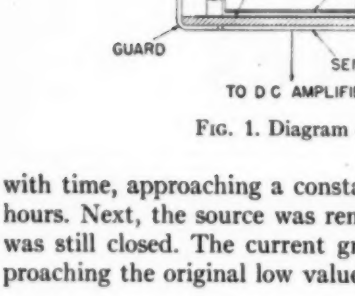


FIG. 2. Effect on conductivity of irradiation. Source: U, $a=0.035$ cm, $V=46$ v. Solid lines calculated from Eqs. (4) and (5).

in Fig. 2. The abscissa is time in minutes. The thickness a of the paraffin was 0.038 cm, and the voltage V, 46 volts. Both the ascending and descending conductivities are shown on the same figure, the circles and crosses being measured values.

In order to interpret curves of the type obtained, it is assumed that the number n of ion pairs as a function of time t is determined by the relation

$$dn/dt=g-\alpha n, \tag{1}$$

in which g is the number of ion pairs generated by the radiation per cc and sec., and α is the removal coefficient given approximately by[4]

$$\alpha=(2mV/a^2)+\gamma, \tag{2}$$

where m is the average ionic mobility and γ is the removal coefficient due to secondary reactions. If conductivities σ are to be calculated, the relation

$$\sigma=2mqn \tag{3}$$

is used (q=charge of an ion; for monovalent ions: 1.6×10^{-19} coulomb). For the ascending σ-curve one then has:

$$\sigma=\sigma_0(1-e^{-\alpha t}), \tag{4}$$

and for the descending σ-curve

$$\sigma=\sigma_0 e^{-\alpha t}, \tag{5}$$

σ_0, the equilibrium conductivity, being related by Eq. (3) to

$$n_0=g/\alpha. \tag{6}$$

The mechanism assumed is simplified in that it disregards space-charge formation and field distortion. However, with fields of 2000 v/cm, and current densities of 10^{-14} amp./cm², a field distortion of only about 5

TABLE I. Evaluation of experimental data from the Uranium Series.

a, cm	V, volt	$(d\sigma/dt)_{t=0}$ 10^{-18}mho/(cm×sec.)	σ_0 10^{-18}mho/cm	α sec.$^{-1}$
0.015	10.5	0.037	5.3	0.0070
0.015	23	0.037	3.4	0.011
0.015	33.5	0.037	2.2	0.017
0.015	46	0.037	1.8	0.020
0.035	23	0.027	8.0	0.0033
0.035	46	0.027	6.0	0.0045
0.035	69	0.027	4.0	0.0067
0.035	92	0.027	3.0	0.0089
0.068	23	0.018	8.3	0.0022
0.068	46	0.018	6.4	0.0029
0.068	69	0.018	5.8	0.0032
0.068	92	0.018	4.0	0.0046
0.068	115	0.018	3.5	0.0052

percent may be produced in 30 minutes. The effect of space-charge on the bombardment conductivity of diamond has recently been investigated by Newton.[4a]

In applying Eqs. (4) and (5) to the data in Fig. 2, it is noted that a particular requirement of Eqs. (4) and (5), namely that the two curves be symmetrical, is fairly well satisfied by the experimental data. In contrast to this, the points do not satisfy a single exponential. It seems that the "up-curve" satisfies a rather steep exponential below $\sigma=6.5\times10^{-18}$, and another flatter one beyond that point. The same holds true for the "down-curve." It is thus assumed that two individual processes are involved: the first, steep portion of the curve is caused by a primary ionization and

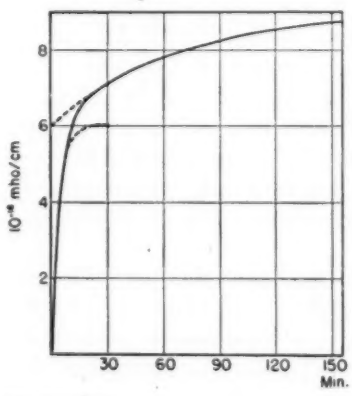


FIG. 3. Calculated conductivity trend based on two components of the type of Eq. (4).

[4] Andrew Gemant, Phys. Rev. 58, 904 (1940).

[4a] R. R. Newton, Phys Rev. 75, 234 (1949).

the second, flat portion by secondary ion-formation from the primary ions. Each part in itself may be described by Eqs. (4) and (5). As an illustration, Fig. 3 shows a curve calculated from the superposition of two single curves of the type of Eq. (4). The sum of the two components has a form similar to the experimental trend shown in Fig. 2, as observed in the majority of the test runs.

Only the first, steep portions of the experimental curves will be evaluated numerically. The results of this computation are summarized in Table I for the uranium source series. The first column lists values of a, the second column lists values of V. The third column lists experimental values of $(d\sigma/dt)_{t=0}$ from the initial slopes (t in seconds). The experimental data permit a determination of these slopes as to order of magnitude only. Actual values were assigned in agreement with their computable ratios. The values of the initial rate of ion generation for a given source must be proportional to the absorbed energy. In a later part of this paper these energies will be calculated. Reference is made to Table IV from which it is seen that the ratios for the three thicknesses a used are 6.7:4.8:3.4. Obviously, the thinner the sample is, the greater will be the specific absorption.

The fourth column of Table I lists the experimental values of σ_0 as determined by the value of σ corresponding to the break in the slope at which the steep portion terminates and the flat portion begins. From the ratio of the data in the third and fourth columns, by virtue of Eq. (6), α in sec.$^{-1}$ is obtained. These values are plotted against V/a^2 in Fig. 4. If our theory is basically correct, the curve, because of Eq. (2), must be a straight line. This is approximately the case.

By using $\sigma_0=6\times10^{-18}$ and $\alpha=0.0053$ for the steep portion, the curves in Fig. 2 were calculated. They reproduce the experimental data fairly well. As another example of the uranium series Fig. 5 is presented, the solid lines showing computed curves as before.

III

The results on the carbon 14 series are presented next. Figure 6 shows the result referring to $a=0.015$ cm and $V=33.5$ volt. In contrast to the uranium series, the two curves are not symmetrical, the rate of the down-curve being markedly less than that of the up-curve. A similar difference, although to a less pronounced degree, is noticeable on some of the U-experiments, in Fig. 2, for instance. One possible reason is that the secondary reaction of the primary ions is not monomolecular but rather bimolecular.

The experimental data are shown in Table II, which is analogous to Table I on the uranium series. Since the soft beta-rays are completely absorbed in the layer of paraffin, the values of g must be inversely proportional to the thicknesses.

In plotting the obtained values of α against V/a^2

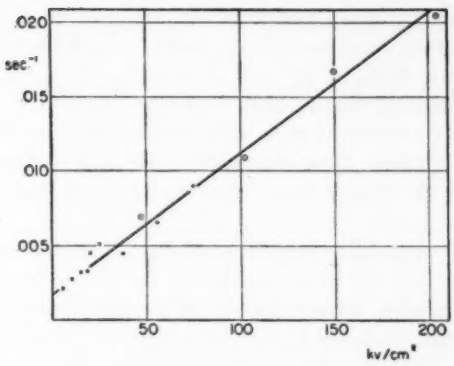


FIG. 4. Removal coefficient α for the uranium series. ○ $a=0.015$ cm, + $a=0.035$ cm, × $a=0.068$ cm.

(Fig. 7), it may be noted that they do not obey a linear relation as closely as in the case of the uranium series. This may be due partly to the circumstance that Eqs. (2) and (6) are only approximately valid for this series.

The radiation effect was sought, besides paraffin, in two polymer materials: polystyrene and tetrafluoro-

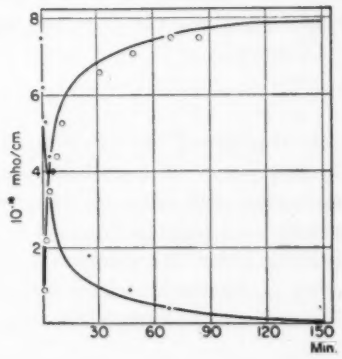


FIG. 5. Source: U, $a=0.015$ cm, $V=10.5$ v. Solid lines calculated from Eqs. (4) and (5).

ethylene. Neither of the two radioactive sources produced measurable changes in the residual currents. These findings help to dispel possible doubts as to whether the observed effect is really caused by an ionization of paraffin wax. In this connection, calculation shows that the number of betas absorbed in the dielectric is entirely negligible as compared with values of g (Table V) in the case of uranium. In C 14 the betas absorbed might amount to 17 percent of g in paraffin and would have to be considered in a more refined interpretation of the data.

IV

The results of this investigation may be used to compute, by means of Eq. (2), the average mobility m of the primary ions in paraffin wax. The slope of the solid line, as drawn through the experimentally obtained values of α, is the same in Figs. 4 and 7, namely, 9.5×10^{-5}, if V is measured in *kv*. From this, $m=4.8\times10^{-8}$ cm/sec. per v/cm, a value that may he considered correct as to its order of magnitude. This value is relatively large and shows that the motion of ions in

paraffin wax is not governed by Stokes' relation. The steady state macroscopic viscosity, as calculable from the plastic flow, is perhaps of the order of 10^{10} poises, from which a much lower mobility than 10^{-8} would follow. The mechanism of ionic displacement is different from a Stokes' type of motion; it may consist, for instance, in jumps from one equilibrium position to another.

If it is assumed that the mobility obtained is characteristic also for ions responsible for the residual current in paraffin, a comparison can be drawn with hydrocarbon oils. The mobility of ions in these latter[4] is very close to, perhaps three to five times larger than 5×10^{-8}. The residual conductivity of such oils is, however, 10^{-15} to 10^{-14} mho/cm as compared with 10^{-19} to 10^{-18} for paraffin wax—a ratio of about 10^4:1. Since the conductivity according to Eq. (3) is proportional to the product of mobility and concentration, it follows that the reduction in conductivity that takes place when a hydrocarbon crystallizes is essentially due to a corresponding reduction in ion concentration.

The intercepts of the lines in Figs. 4 and 7 give the value of γ, the rate coefficient of secondary reactions—0.0017 sec.$^{-1}$ for the U-series, and 0.0033 sec.$^{-1}$ for the C 14 series.

V

By combining the experimentally obtained values of $(d\sigma/dt)_{t=0}$ with the mobility m, the efficiency of ion generation may be estimated. From Eq. (3) the rate of primary ion-pair production g is calculated and, by dividing it by the electron energy absorbed, the efficiency is obtained, at least as to its order of magnitude.

In order to be certain that the effect is caused essentially by the 2.32-Mev electrons in the U-series, a few filter experiments were carried out, the results of some of which are shown in Table III. The table gives in its last column the increase in σ during the first fifteen minutes. The transmission was calculated from the absorption coefficient 6.2 cm^2/g, valid for the 2.32 Mev electrons. The change in σ is approximately proportional to the intensity, as can be seen. (Strictly speaking, proportionality with the energy transmission would be a better criterion.) For completeness' sake a test on C 14 is also included (absorption coefficient 270 cm^2/g).

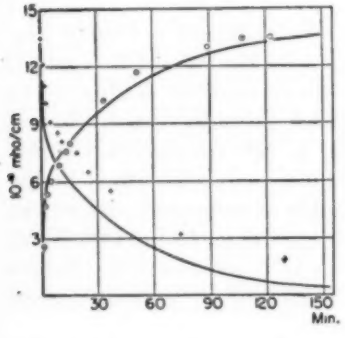


FIG. 6. Source: C 14, a=0.015 cm, V=33.5 v. Solid lines calculated from Eqs. (4) and (5).

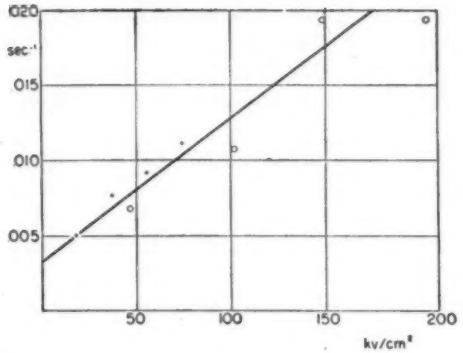


FIG. 7. Removal coefficient α for the carbon 14 series. ○ a=0.015 cm, + a=0.035 cm.

With this information the energy absorbed for the two sources and three thicknesses can be calculated; the results are given in Table IV. The particle output was measured by a Geiger counter, as mentioned in the first section of this paper; multiplying these figures by the average energies (0.77 Mev for U, 0.048 Mev for C 14) the energy output is obtained. Next the particle absorption is computed for various values of a. The energy absorption is then obtained graphically from the known energy distribution curve of betas;[5,6] this evaluation is approximate only. By multiplying the figures in Columns 3 and 6, and dividing by the values of a, the last figure, namely the specific absorption, is obtained for each case.

The efficiency of ionization can now be computed, and the result is summarized in Table V. The figures are based on data for a=0.015 cm. The values of g are

TABLE II. Evaluation of experimental data from the Carbon 14 Series

a, cm	V, volt	$(d\sigma/dt)_{t=0}$ 10^{-18}mho/(cm×sec.)	σ_0 10^{-18}mho/cm	α sec.$^{-1}$
0.015	10.5	0.012	17	0.0069
0.015	23	0.012	11	0.011
0.015	33.5	0.012	6.0	0.019
0.015	46	0.012	6.0	0.019
0.035	23	0.050	10	0.0050
0.035	46	0.050	6.5	0.0077
0.035	69	0.050	5.5	0.0091
0.035	92	0.050	4.5	0.011

TABLE III. Identification of the electrons producing ionization.

Source	a, cm	V, volt	Filter	Betas transmitted, %	$\Delta\sigma$ in 15 min. 10^{-18}mho/cm
U	0.068	23	—	100	11
U	0.068	23	0.11 g/cm^2 Al	50	5.1
U	0.068	23	1.2 g/cm^2 Pb	0	0
U	0.068	92	—	100	4.8
U	0.068	92	0.11 g/cm^2 Al	50	2.6
C 14	0.035	92	—	100	5.5
C 14	0.035	92	0.0024 g/cm^2 Al	50	2.7

[5] Kai Siegbahn, Phys. Rev. **70**, 127 (1946).
[6] A. Gemant and H. M. Archer, Nucleonics **3**, 40 (1948).

TABLE IV. Calculation of electron energies absorbed by paraffin.

Source	Intensity: Particles/ (cm²×sec.)	Intensity: 10³Mev/ (cm²×sec.)	a, cm	Absorption: Particles, %	Absorption: Energy, %	Absorption: 10³Mev/ (cm³×sec.)
U	500	0.38	0.015	7.9	26	6.7
U	500	0.38	0.035	17	43	4.8
U	500	0.38	0.068	31	60	3.4
C 14	20,000	0.96	0.015	98	100	64
C 14	20,000	0.96	0.035	100	100	27

TABLE V. Calculation of efficiency of primary ionization ($a=0.015$ cm).

Source	$(d\sigma/dt)_{t=0}$ 10^{-18}mho/ (cm×sec.)	g cm^{-3} sec.$^{-1}$	Absorption 10^3 Mev/ (cm³×sec.)	Ion pairs/ Mev
U	0.037	2.4×10^6	6.7	360
C 14	0.12	7.9×10^6	64	125

obtained from $(d\sigma/dt)_{t=0}$ according to Eq. (3), i.e., division by 15.3×10^{-27}. The figures in the last column are rather low if compared with 20,000, the corresponding order of magnitude in gases. One must realize that only ions of a mean life of one minute or more were registered by the experiments described here; they are probably a small fraction of all ions produced. Only small radicals, such as H^+, H^-, CH^+, CH^-, and the like, have a sufficiently high mobility to be able to move away from the heavy and relatively immobile counter-ions and escape immediate recombination. Only the electrons colliding with the end of chain molecules produce such types of ions. This explanation is also in keeping with observations on the effect of admixtures to paraffin and observations on polymers.

VI

The polarity of the voltage in all experiments described was such that the graphite electrode was positive. If the polarity is reversed, the effect remains essentially the same, but its magnitude is increased with the C 14 source. This is borne out by Table VI.

The polarity effect with C 14 might be caused by the fact that ionization is more pronounced near the top of the paraffin layer than further below, producing a higher conductivity near the top.

Next, the influence of admixtures to paraffin wax upon the radiation effect was investigated. It appears

TABLE VI. Effect of the polarity of voltage.

Source	a, cm	V, volt	Polarity of top electrode	$\Delta\sigma$ in 10 min. 10^{-18} mho/cm
U	0.035	23	+	7.5
U	0.035	23	−	6.4
U	0.015	23	+	3.4
U	0.015	23	−	2.6
C 14	0.035	23	+	10
C 14	0.035	23	−	22
C 14	0.015	23	+	11.5
C 14	0.015	23	−	16.5
C 14	0.035	92	+	5.1
C 14	0.035	92	−	7.0

that they reduce the magnitude of the effect. As an example, paraffin with five percent myristic acid is quoted. With C 14, $a=0.015$ cm and $V=23$ v, the 10 min. increase in σ was 8.6×10^{-18} mho/cm, as compared with 11.5×10^{-18} in pure paraffin wax. Simultaneously, the residual conductivity without irradiation increased from 0.8×10^{-18} to 14×10^{-18}. In beeswax, which contains various acids and esters, no radiation effect was found. Similarly, if the paraffin layer was produced in a brass cell, instead of a stainless steel cell, the effect was reduced, probably because of the presence of oxidation products in the wax. With C 14, $a=0.015$ cm and $V=46$ v, the change of σ in 70 minutes was 7.9×10^{-18} mho/cm as compared with 12.8×10^{-18} in the stainless steel cell. The reduction of the effect in all these instances may be explained by an increase of the coefficient γ, since the presence of acids and esters promotes secondary reactions. This shows the degree of instability of the ions in question, in keeping with the low efficiency figures found.

As mentioned in Section III, no effect was found in polymers; this shows that ionization is much less intensive in a macromolecular system with long chain molecules than in a solid hydrocarbon of medium molecular weight. This is in keeping with the assumption that only small charged radicals that split off the ends of a chain molecule have sufficient mobility to move away from the point of their origin and are detectable by means of their electrical conductivity.

Part of the measurements was carried out by E. L. Alexanderson, and the tridecanoic acid was prepared by Edward Hines; both of the Research Department of The Detroit Edison Company.

Antennas and Open-Wire Lines. Part II: Measurements on Two-Wire Lines*

K. Tomiyasu†
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received December 20, 1948)

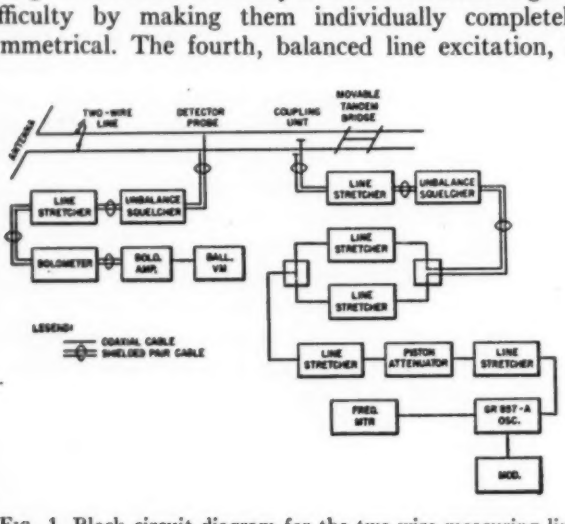
The open two-wire line is investigated as a measuring instrument and difficulties in its operation are overcome. These include, in particular, the problem of balance. The impedance of antennas is shown to depend on the nature of the driving structures. A comparison of measured impedances and corresponding theoretical values shows good agreement.

1. INTRODUCTION

A TWO-WIRE line, being a balanced structure, is appropriate to drive and to measure impedances of symmetrical dipole antennas. For this purpose a properly designed measuring line and a very sensitive standing-wave detector are required. In this paper, the design and construction of a two-wire measuring line is described.

2. METHOD OF OBTAINING A BALANCED SYSTEM

The use of two-wire lines has been limited by the belief that such lines necessarily radiate excessively. Actually, radiation from a closely spaced, *balanced* line is very weak.[1] Strong radiation is due to unbalanced currents, i.e., to currents on the two lines that are not of equal magnitude and not of opposite phase.[2] Such currents can be prevented by independently balancing the load, the two-wire line, the standing-wave detector, and the line excitation. These are shown schematically in Fig. 1. The first three may be balanced without great difficulty by making them individually completely symmetrical. The fourth, balanced line excitation, is more difficult to obtain. This agrees with the experience of other investigators.[3]



* Fig. 1. Block circuit diagram for the two-wire measuring line. A shielded-pair bolometer standing-wave detector is used. The frequency was 300 mc/sec.

Since most commercial oscillators have a coaxial output, it is necessary to use a coaxial-to-shielded pair transformer that does not introduce unbalanced components. The two-wire line can then be excited from the open end of the shielded-pair line.

Since the unbalanced component of two unequal currents treats a shielded-pair line as a coaxial line, it can be stopped by using a suitable bridge called an "unbalance squelcher."[4] This is the double stub device shown schematically in Fig. 2. Each stub is a half-wave-length long and has a short-circuiting bar between the two inner conductors a quarter-wave-length from the junction. The two stubs are a quarter-wave-length apart and form a short-circuiting bridge for the unbalanced component. The balanced component is not affected, since only high impedance stubs are connected across the line for this component.

The unbalance squelcher was tested in the shielded-pair circuit in conjunction with several devices for obtaining balanced excitation on two-wire lines and was found to give perfect balances for all.[5] The circuit actually used for transforming from the coaxial line to the shielded pair consisted of a coaxial "T" junction with one line longer by a half-wave-length as shown schematically in Fig. 1.

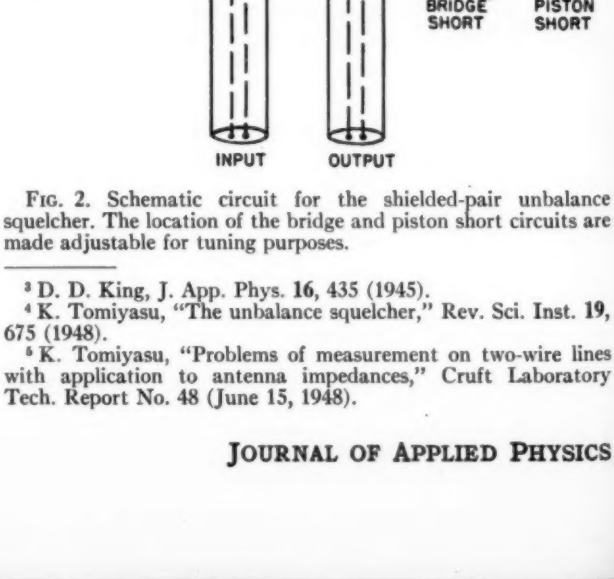


Fig. 2. Schematic circuit for the shielded-pair unbalance squelcher. The location of the bridge and piston short circuits are made adjustable for tuning purposes.

* Presented at Joint Meeting U.R.S.I. and I.R.E., Washington, D. C., May 4, 1948. This research was supported in part by the Navy Department (ONR) and the Signal Corps, U. S. Army, under Contract N5ori-76, T.O. 1.

† Now at Sperry Gyroscope Company, Great Neck, New York.

[1] R. King, *Electromagnetic Engineering* (McGraw-Hill Book Company, Inc., New York, 1945), Vol. I, Chapter VI.

[2] R. King, Proc. I.R.E. **31**, No. 11 (November, 1943).

[3] D. D. King, J. App. Phys. **16**, 435 (1945).

[4] K. Tomiyasu, "The unbalance squelcher," Rev. Sci. Inst. **19**, 675 (1948).

[5] K. Tomiyasu, "Problems of measurement on two-wire lines with application to antenna impedances," Cruft Laboratory Tech. Report No. 48 (June 15, 1948).

3. THE TWO-WIRE LINE

An important factor in designing a two-wire line for measuring antenna impedances is its length. A short line places the equipment and operator so close to the antenna that reflections from them may alter the measured impedance. A very long line, on the other hand, introduces excessive line attenuation and the problem of supports. In the apparatus actually used, a two-wire line 35 feet long was supported by tension. At an operating frequency of 300 megacycles per second the equipment and operator were sufficiently far from the antenna to prevent serious interaction. The copper conductors of the line had a diameter of 0.120 in.; they were spaced 2.0 cm to give a characteristic impedance of 308 ohms, a value particularly chosen to yield greatest accuracy in all antenna impedance measurements. The attenuation constant was 1.55×10^{-3} neper per meter, which was low but not negligible for the long line.

4. THE GENERATOR

A 300-megacycle per second r-f signal from a General Radio 857-A butterfly oscillator was tone-modulated by a Hewlett Packard audio oscillator in order to use a bolometer detector. The output was calibrated by a coaxial piston attenuator, transformed to a shielded pair, and passed through the unbalance squelcher and line stretcher to the coupling unit. The generator and modulator were placed in a steel box to reduce stray fields.

In addition to short-circuiting the unbalanced r-f components, the unbalance squelcher presents a zero impedance to audio frequency voltages and, thereby, prevents them from appearing at the coupling unit. The audio frequency signals before being checked in this manner radiated across distances as great as four feet and made the detector inoperative.

5. SMALL-LOOP CRYSTAL PROBE

The types of standing-wave detectors previously used on two-wire lines have been simple crystal probes using resonant sections of two-wire transmission line. In order to provide for a better impedance match for the crystal detector and to obtain a definite and small area of interaction with the electromagnetic fields, a section of line was shortened to a small loop and made resonant with a series condenser. This detector, called a small-loop probe, is shown schematically in Fig. 3. Tuning is accomplished by varying the area of the loop. A 1N21-B crystal is in parallel with the condenser. An r-f by-pass condenser provides an open circuit for d.c. The rectified current activates a remote meter through leads made of nichrome wire to attenuate r-f currents. The small loop probe can be used as either a voltage or a current detector by suitable orientation relative to the two-wire line. It is mounted on a movable base.

Although the small loop probe was quite sensitive and satisfactory for qualitative measurements, its effect on the line was not negligible. A detector having

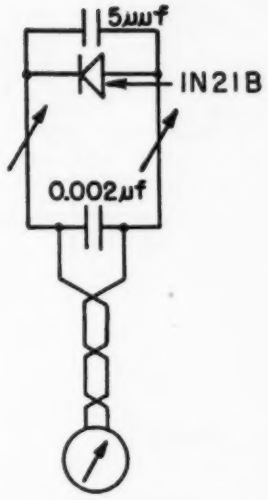


FIG. 3. Schematic circuit for a small loop crystal probe for use on a two-wire line. A very sensitive microammeter is required.

a finite shunting admittance across a transmission line can introduce significant errors in the impedance measurements. For a non-zero admittance, the apparent standing wave ratios are smaller than, and have definite relationships to, the true standing wave ratios.[5,6] If the admittance has an imaginary component,

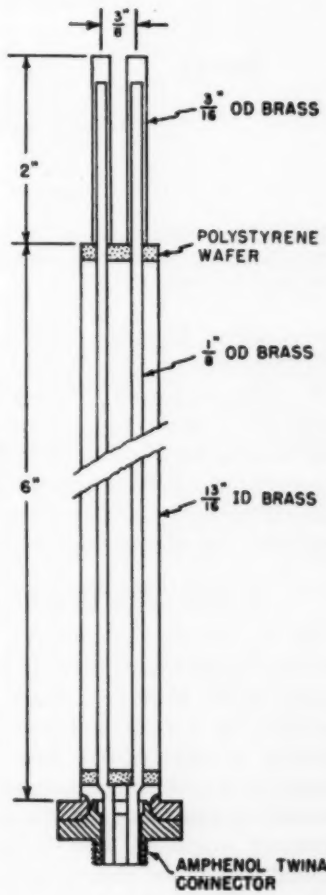


FIG. 4. Cross-sectional view of the shielded-pair probe. The probe receives its excitation from the extended central conductors.

[6] C. G. Montgomery, *Technique of Microwave Measurements* (McGraw-Hill Book Company, Inc., New York, 1947), Radiation Laboratory Series, Chapter 8.

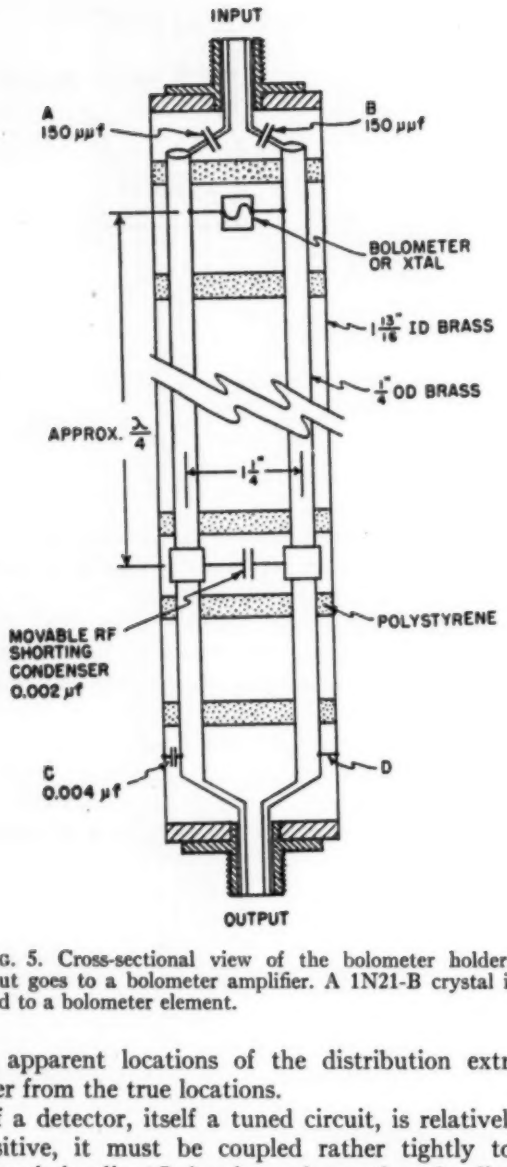


FIG. 5. Cross-sectional view of the bolometer holder. The output goes to a bolometer amplifier. A 1N21-B crystal is preferred to a bolometer element.

the apparent locations of the distribution extremes differ from the true locations.

If a detector, itself a tuned circuit, is relatively insensitive, it must be coupled rather tightly to the transmission line. It has been shown that the distribution curve of a detector coupled greater than critical is characterized by a double-hump curve.[5,7] This leads to inaccurate impedance measurements.

6. THE SHIELDED-PAIR DETECTOR

The sensitivities of detectors used on coaxial lines and wave guides have been made so high that detector loading is less than line attenuation. These detectors, operating on a tone-modulated signal, use bolometer elements. A shielded-pair line was used as the basic element in a new standing wave detector consisting of the shielded-pair probe, unbalance squelcher, bolometer, bolometer amplifier, and Ballantine voltmeter. The shielded-pair probe shown in Fig. 4 is made of brass tubings and the probe tips are extensions of the inner conductors. The outer shield has an inside diameter of $\frac{13}{16}$ in. and the two inner conductors are $\frac{1}{8}$ in. in diameter and spaced at $\frac{3}{8}$ in. between centers. The characteristic impedance is 163 ohms. The probe has an Amphenol Twinax connector at its base for the Amphenol 81-18B shielded pair cable having a characteristic impedance of 160 ohms. (The coupling unit for the generator circuit is almost identical with this probe.) In Fig. 5, the bolometer holder is shown. A bolometer is shown as the detecting element, but a 1N21-B crystal can be used with excellent results. The two condensers *A* and *B* were necessary to remove from the bolometer the d.c. short circuit in the unbalance squelcher. The distance between the bolometer and the r-f short-circuiting condenser is about one-quarter wave-length to terminate effectively the r-f line at the bolometer. Short-circuit *D* was necessary to place one point of the input circuit of the bolometer amplifier at ground potential. Condenser *C* by-passed undesired r-f signals. The modulated signal is dissipated in the detecting element and the audio frequency signal derived from the element is amplified by a bolometer amplifier. This amplifier is in a steel box to reduce interaction with stray fields. It is tuned at 1040 cycles, has a band width of 3 percent, a gain of 123 db on the same impedance level, and is almost linear for a 60-db range in input signal.

7. AUDIO FREQUENCY POWER FROM AN AMPLITUDE-MODULATED SIGNAL

From a theoretical analysis, it can be shown that the audiofrequency power obtained from an amplitude-modulated signal has two frequency components separated by one octave.[5] When both modulator and bolometer amplifier are tuned to the same frequency, the detector output,

$$e_d \propto E_0 E_s.$$

When the modulator frequency is one-half the fre-

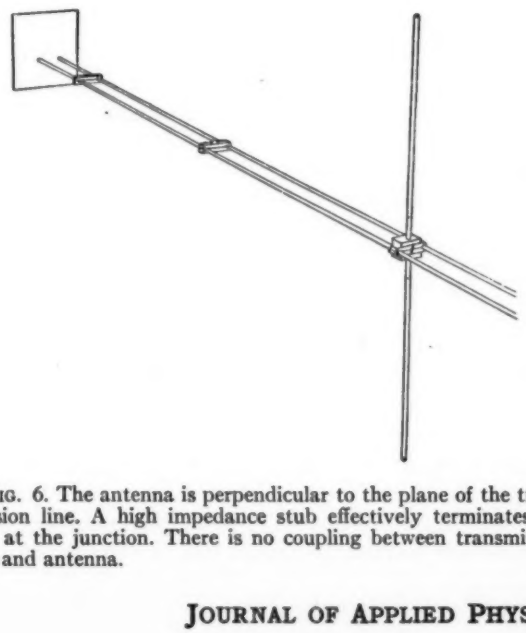


FIG. 6. The antenna is perpendicular to the plane of the transmission line. A high impedance stub effectively terminates the line at the junction. There is no coupling between transmission line and antenna.

[7] K. Tomiyasu, "Loading and coupling effects of standing-wave detectors," submitted for publication to I.R.E.

quency of the tuned bolometer amplifier,

$$e_d \propto E_s^2/4,$$

where E_0 is the amplitude of the unmodulated r-f signal, and E_s is the amplitude of the modulating signal. The output in the former case is four times that in the latter case for 100 percent modulation, and much greater than four times for smaller degrees of modulation. The preferred condition is with the modulating frequency equal to the tuned frequency of the bolometer amplifier. Evidently, the quantity E_0 can be determined by taking the ratio of the two outputs.

8. CALIBRATION OF THE DETECTOR

A precise calibration of the detector was found to be rather difficult, but a satisfactory result was obtained using an r-f piston attenuator as the standard in the generator circuit. A Littelfuse bolometer element for the detector had a square-law characteristic but was relatively insensitive. A 1N21-B crystal, although more sensitive by 30 db, had a calibration curve which was almost linear for low degrees of modulation and became square-law when the modulation exceeded 10 percent. All antenna impedance measurements were made with a modulation of 45 percent. Since the bolometer amplifier was slightly non-linear, all deflections had to be corrected.

The purpose of the unbalance squelcher in the detector circuit was threefold. It suppressed unbalanced r-f signals, it prevented stray audio fields from entering through the probe tips, and it made possible the use of the more sensitive but electrically unbalanced 1N21-B crystal.

9. ANTENNAS AND THEIR DRIVING STRUCTURES

The first antenna to be measured was placed perpendicular to the plane of the transmission line as shown in Fig. 6, with a high impedance stub beyond the antenna to terminate the line effectively at the junction. The stub was necessary to support the wires. The driving structure has a small gap 0.022 in. wide milled at an angle of 45 degrees. See Fig. 7. Each half-bridge is made of $\frac{1}{4}$-in. square brass and is clamped to the two-wire line. The antennas are $\frac{1}{8}$ in. in diameter and made of short, threaded, brass sections having lengths of 0.1, 0.2, 0.5, and 1.0 radian. A small polystyrene frame aligned the half-bridges, kept the line spacing and gap width constant, and maintained the collinearity of the antenna.

In Part I it has been shown that there is no mutual coupling between antenna and transmission line for this type of driving structure with the antenna perpendicular to and in the neutral plane of the line. There is, however, a reactive terminal-zone effect due to the discontinuous nature of the transmission line and the residual capacitance of the driving structure. Instead of compensating for this reactance mathematically, it was tuned out electrically. Thus, the stub beyond the junction was made about one-eighth of a wave-length long in order to place a voltage maximum at the junction with the antenna removed.

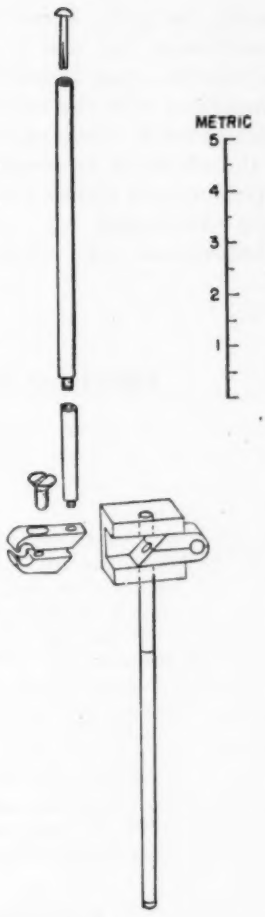


FIG. 7. An exploded view of the driving structure for the antenna shown in Fig. 6.

With this structure, antenna impedances were measured and the good agreement with the King-Middleton second-order theory (Fig. 30 of Part I) indicates that a slice generator has been approximated, using a two-wire-line excitation.

The second antenna to be measured was oriented in the plane of the transmission line as shown in Fig. 8. A quarter-wave-length stub placed a voltage maximum at the junction. It has been shown that there is mutual coupling between antenna and transmission line in-

FIG. 8. The antenna is in the plane of the transmission line.

cluding the stub. There is also an end-effect on the transmission line and a gap-effect on the antenna. Impedances were measured for this antenna and a comparison with the corrected theory (Figs. 24a and 24b in Part I) simultaneously confirms the major role of the effects of the terminal zone upon the antenna impedance and verifies the analytical method of taking them into account.

Impedances were measured by the standing-wave-ratio and the resonance curve peak-width methods.[8] All measurements had to be corrected for the finite attenuation due to the long measuring line.

ACKNOWLEDGMENT

The author wishes to express his appreciation to Professor R. W. P. King who suggested the problem and gave valuable advice.

[8] D. D. King, Proc. I.R.E. **35**, 509–514 (1947).

Effects of Stress and Deformation on the Martensite Transformation*

ANDREW W. MCREYNOLDS**
Institute for the Study of Metals, University of Chicago, Chicago, Illinois

(Received January 31, 1949)

The relation between elastic and plastic strains and the martensite transformation from face-centered to body-centered structure has been investigated in Fe-Ni 71–29 alloy. Transformation begins at about −30°C and is of the martensite type except that it is not all completed instantaneously but continues at decreasing rate at constant temperature.

It was found that (1) There is no anomalous decrease in elastic moduli as the M_s temperature is approached, nor is M_s affected by applied elastic stresses. (2) Plastic deformation in a temperature range above M_s causes partial transformation in amounts decreasing to zero at temperature $M_d \approx M_s + 75$°C, but stabilizes against further transformation by cooling. (3) Plastic yield stress increases sharply from M_d to M_s then drops suddenly to a low value just below M_s. (4) The M_s temperature depends on grain size; a factor of ten increase in grain size reduces M_s about 30°C.

These results are contrary to expectations on the basis of the usual homogeneous shear model for the transformation mechanism. They indicate that the martensite phase becomes thermodynamically stable at M_d but that transformation does not begin until M_s because of surface and strain energy considerations in growth of a martensite plate from a small nucleus.

I. INTRODUCTION

PHASE transformations in solids are observed to occur by two relatively distinct mechanisms, one proceeding by the growth at continuous rate of a nucleus of the new phase and the other discontinuously by the almost instantaneous appearance of definite regions of the new phase, by a sudden, diffusionless lattice shear. The former reaction depends on diffusion rates and may proceed slowly, but in the latter type of transformation the time required for formation of each region is of the order of the time for propagation of an elastic wave.

The best known example of this diffusionless mechanism is the martensite reaction in steel. When slowly cooled, steel changes from face-centered cubic to body-centered cubic by a growth process involving diffusion of carbon, but on rapid cooling the change takes place by the diffusionless martensite reaction. The principal characteristics of the martensite reaction are:

(1) As austenite is cooled, transformation begins at a rather sharply defined temperature (the M_s point).

(2) It does not progress at constant temperature, but only while further cooling continues.

(3) Consequently, the amount transformed by cooling to a fixed temperature is independent (within limits) of the cooling rate, and depends only on chemical composition and grain structure.

(4) Martensite forms by the sudden (less than 0.001 second) appearance of plate-shaped regions within each grain.

(5) There is a characteristic orientation of the plates and also of the crystallographic axes of the martensite (b.c.c.) relative to the austenite (f.c.c.).

Although the term martensite originally referred to the hard body-centered tetragonal structure in iron-carbon alloys, similar reactions have been observed in many other pure metals and alloys such as iron-nickel,[1–3] iron-manganese,[4] copper-aluminum,[5] copper-zinc,[6] lithium,[7] lithium-magnesium,[7] cobalt,[8] and zirconium.[9] In accordance with the suggestion of Greninger and Troiano,[10] who pointed out the similarity of these reac-

* This research has been supported in part by the ONR contract n-6 ori-20-IV.

** Now at Brookhaven National Laboratory, Upton, Long Island, New York.

[1] G. Kurdjumow and G. Sachs, Zeits. f. Physik **64**, 325 (1930).
[2] Z. Nishiyama, Sci. Rep. Tohoku Imp. Univ. **I, 23**, 638 (1934–35).
[3] A. B. Greninger and A. R. Troiano, Trans. A.I.M.E. **145**, 91 (1941).
[4] A. R. Troiano and F. T. McGuire, Trans. A.S.M.E. **31**, 340 (1943).
[5] A. B. Greninger, Trans. A.I.M.E. **133**, 204 (1939).
[6] A. B. Greninger and V. G. Mooradian, Trans. A.I.M.E. **128**, 337 (1938).
[7] C. S. Barrett and O. R. Trautz, Trans. A.I.M.E. **15**, 1 (1948).
[8] A. R. Troiano, Met. Tech. T P 2348 (April 1948).
[9] W. G. Burgers, Physica **1**, 561 (1934).
[10] A. R. Troiano and A. B. Greninger, Metal Progress **50**, 303 (1946).

tions, the transformations in iron-nickel are referred to herein as martensite.

Very little is known about the anomalies in elastic and plastic behavior at temperatures near or in the transformation range. Such information is of interest, partly as empirical data, but particularly to reveal the fundamental mechanism of transformation. In spite of numerous experiments on martensite from different viewpoints, neither the atomic motion by which a plate of the new phase is formed nor the physical conditions which cause and control transformation is understood. The particular questions of interest are whether the change of lattice occurs by a homogeneous shear or by a more complex rearrangement of adjacent atoms, and what changes in the lattice occur as temperature decreases toward the M_s point, which render the face-centered structure unstable and allow transformation to body-centered. Since a study of the relation of elastic and plastic properties near M_s gave promise of partly answering these questions, the present studies were undertaken. Iron-nickel alloy was chosen as a convenient typical case.

Many studies of the lattice orientation relationship of martensite to austenite have been made on iron-carbon[1,11] and iron-nickel[2,11] alloys. Although they differ in detail, the Kurdjumow-Sachs relationship,

$$(111)\gamma \| (011)\alpha, \quad [101]\gamma \| [111]\alpha,$$

the Nishiyama relations:

$$(111)\gamma \| (011)\gamma, \quad [112]\gamma \| [011]\alpha,$$

and the Greninger-Troiano relations (between the above two) all lie rather close together and all were explained on the basis of single or double homogeneous shears*** of the lattice, followed by dimensional readjustment. Recently, Barrett[12] proposed that the lattice change could also be effected by a heterogeneous shear, possibly by passage of a group of dislocations along adjacent planes causing small shears of successive layers in various directions.

In a transformation proceeding by homogeneous shear both elastic stresses and plastic strains would be expected to play an important role. Although transformation by plastic deformation has been observed qualitatively in a few cases, practically no quantitative investigations of the conditions of this phenomenon have been made. Strain transformations of the martensite type were found by Greninger and Mooradian[6] in 60:40 beta-copper zinc, and by Scheil[13] in iron-nickel,

[11] G. Wasserman, Mitt d. Kais.-Wilhelm Inst. f. Eisenf. **17**, 149 (1935).

*** Kurdyumow and Sachs proposed that transformation occurs by a shear about 15° on the (111)γ-plane in the [211]γ-direction followed by a second shear on the (211)γ-plane. Nishiyama proposed a single shear of 19° 28′, also on the (111)γ-plane in the [211]γ-direction. Greninger and Troiano postulated a shear on the irrational habit plane of the plate then on the (112)α-plane in the (111)a direction.

[12] C. S. Barrett, private communication.

[13] E. Scheil, Zeits f. anorg. u. allgem. Chemie **207**, 21 (1932).

TABLE I.

Alloying elements	Alloy *I* (%)	Alloy *II* (%)
Nickel	28.74	29.41
Manganese	0.39	0.32
Carbon	0.01	0.014

the effect being limited in the latter case to a range of temperatures above the M_s point. Of particular interest is the recent work of Barrett[7] in this laboratory, showing a martensite reaction at liquid nitrogen temperatures in lithium and lithium-magnesium alloys, with deformation-transformation properties paralleling those in iron-nickel found by Scheil[13] and by the author. Here also transformation occurred only in a limited temperature range above M_s. In lithium, however, the structure produced by deformation differs from that produced by cooling in that errors in stacking close-parked planes over each other are common in the spontaneous product, but are few in the strain-induced product. Deformation-transformation between different types of martensite structures was observed in copper-aluminum by Greninger[5] and in iron-maganese by Troiano and McGuire.[4] The only investigation of the effects of stress within the elastic limit apparently is that of Scheil,[13] in which a small effect was observed with large-grained samples but none with fine grains.

From theoretical considerations based on postulated

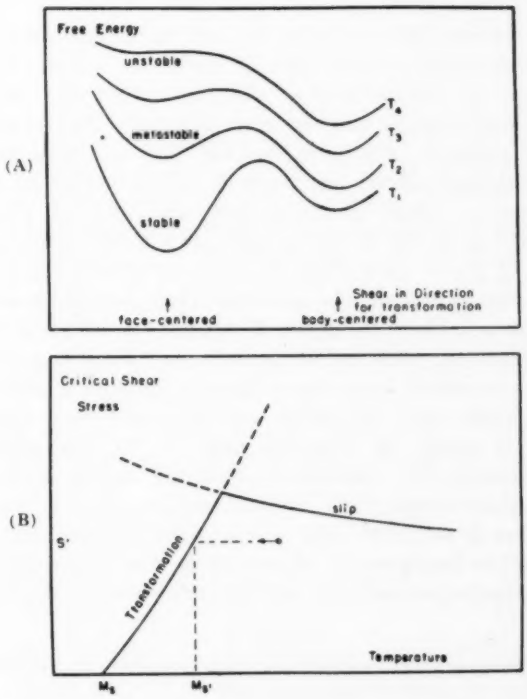
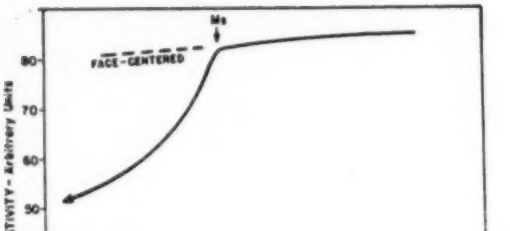


FIG. 1. (A) Hypothetical change of shape of volume free energy curves as temperature decreases from T_1 to T_4 making face-centered structure unstable. (B) Scheils (see reference 13) proposal of slip and transformation as alternative ways of plastic shear motion. Lattice under shearing stress S' should begin to transform at M_s' instead of M_s.



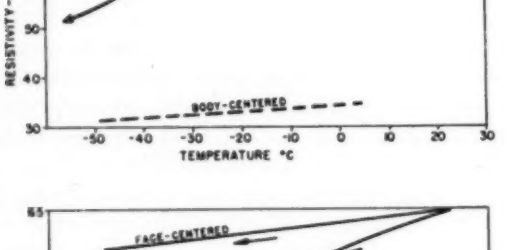


FIG. 2. Electrical resistivity changes during transformation of Fe-29 percent Ni alloy between face-centered and body-centered structure, (above) during continuous cooling, (below) on heating and subsequent cooling, illustrating how resistivity can be used to measure degree of transformation.

mechanisms of transformation, one can make some predictions as to the relations of elastic stress and plastic strain to transformation. Assuming that the lattice motion is a single homogeneous shear with simultaneous readjustment of atomic spacing (like the Nishiyama mechanism) Zener has suggested that a plot of free energy *vs.* shear in the appropriate direction should appear as in Fig. 1. At successively lower temperatures $T_1, T_2, T_3, \cdots$, the shape of the curve could be assumed to vary as shown, the austenite changing from stable to metastable and, at the M_s point, to unstable. The height of the potential hump between the minima represents a critical shear stress for transformation and the curvature near a minimum represents the shear modulus. It might be expected that as the temperature approaches M_s both the macroscopic elastic modulus and macroscopic yield stress would decrease. This prediction is in accord with the properties postulated by Scheil to interpret his observations. He proposed that transformation and slip are alternative ways of plastic strain of the lattice, the critical shearing stresses for the two processes varying as shown in Fig. 1. Thus deformation by transformation would occur spontaneously without stress at M_s, and more readily than slip at any temperatures below the intersection of the two curves. This picture leads to the expectations that the austenite would be very soft within the transformation range, and that the elastic limit would increase from zero at M_s to a maximum at the intersection of the two curves, indicated by M_d in Fig. 1. Plastic strain would occur by transformation between M_s and M_d but by slip above the temperature M_d, which, therefore, represents the approximate upper limit for martensite formation by deformation. Also an applied stress such as S' in Fig. 1 should cause transformation on cooling to begin at a temperature $M_s' > M_s$. Scheil's[13] results on iron −29 percent nickel showed that transformation by deformation does occur within a temperature range above M_s, and also gave indications of decrease of elastic modulus as temperature decreases to M_s, and of slight increase of M_s with applied stress. The latter effect was found only in large-grained crystals. He also found a dependence of the M_s temperature on grain size and suggested that the limitations of the grain boundaries on growth of martensite plates partly determines M_s and also that formation of the first plates further subdivides the grains, necessitating further decrease of temperature to form additional plates.

The specific purpose of the author's investigation was, then, to compare observed mechanical properties near the transformation range with those predicted on the basis of various possible mechanisms of transformation. For this purpose it was necessary to make measurements of anomalies in elastic modulus and elastic limit near the M_s point, effect of elastic stress and plastic strain on the progress of transformation, and the conditions under which transformation could be induced by deformation.

TABLE II.

Sample No.	Grain diam. (mm) (linear count)	M_s temperature
1	0.030	−50°C
2	0.26	−17°
1	0.24	−17°
2	0.028	−45°

II. APPARATUS AND MATERIALS

Alloys

The original ingots were made by melting puron iron with high purity nickel in an atmosphere of purified argon. A small amount of manganese was added in order to increase ductility for drawing operations. After annealing 6 hours at 1300°C for homogenization, the ingot was alternately vacuum annealed at 900°C and swagged, to wire of 0.040-inch diameter. Chemical analyses on the two ingots used showed a variation of less than 0.25 percent in nickel content of top and bottom of the ingot. Average analyses are given in Table I. In most tests the sample was converted to austenite before test by annealing ½ hour at 900°C in vacuum, with resultant grain size of about 34 grains/mm. By annealing 6 hours at 1300°C in vacuum some samples of size 3.8 grains/mm were obtained. Grain sizes here represent the average number of grains per mm intersected by a straight line through the sample.

Apparatus

1. Electrical measurements—Determination of the amount of transformation was made in all cases by the elestrical resistivity method.[14] A continuous record was made in each test by passing a constant current from a storage battery through the sample and recording on a Brown self-balancing potentiometer the potential difference between two leads soldered to different points of the sample wire. Usually only comparative measurement during different stages of test was required.

2. Temperature control was accomplished by immersion in an alcohol bath. Where only cooling was required the bath was agitated by a motor stirrer and cooled by conduction through a copper bar extending from the bath into liquid nitrogen. For compression and some other tests the bath was continuously circulated by a pump through cooling coils immersed in liquid nitrogen. Temperature gradients and fluctuations were within 0.1°C in the former apparatus and within 1°C in the latter.

3. Compression was accomplished between two hardened and ground steel plates in a hydraulic press. The sample, 0.040 in. diameter by 1.7 in. long, was soldered to two spring clips which served as current leads, and potential leads were also soldered on just outside the compressed area. Both blocks and the sample were immersed in the circulating alcohol coolant. A dial gauge device was attached to measure thickness of the compressed sample by the spacing between the compression plates and applied force was measured by a Bourdon gauge on the press. Resistivity measurements were made between compressions, when the sample wire was separated from the metal compression plates. The deformation was introduced by compression because in this way large deformations could be made with practically no change of length and cross-sectional area.

4. Tensile strain apparatus—A special small tensile machine was constructed for measurement of both elongation and transformation under stress. The ends of the three-inch wire samples were soldered into hemispherical end blocks which rested in cone seats, to allow self-alignment. The stress was applied through a lever system by either fixed weights or water flowing into a container. Design was such that the entire sample assembly, surrounded by a brass tube, could be immersed in a circulated liquid bath to vary temperature or hold it at a fixed valve as desired. Strains were measured to a precision of 5×10^{-6} by either an optical lever device or a dial gauge measuring separation of the cone seats. Current and potential leads were attached to each end of the specimen for simultaneous electrical measurement of transformation.

5. Static measurement of rigidity modulus—A six-inch length of the specimen wire, surrounded by a glass tube through which alcohol was circulated, served as suspension for the coil of a modified permanent magnet galvanometer, with motion damped by a vane immersed in oil. Torque was periodically applied simply by passing current through the coils and rigidity modulus measured by the resulting deflection of the light spot.

6. Dynamic measurement of elastic moduli—The top of the specimen was attached to a fixed support and a weight assembly attached to the bottom end. To measure rigidity modulus the period of torsional oscillation of the system was long enough (about 0.5 sec.) to be timed by a stopwatch. Young's modulus was measured by exciting longitudinal vibration by a magnetic drive

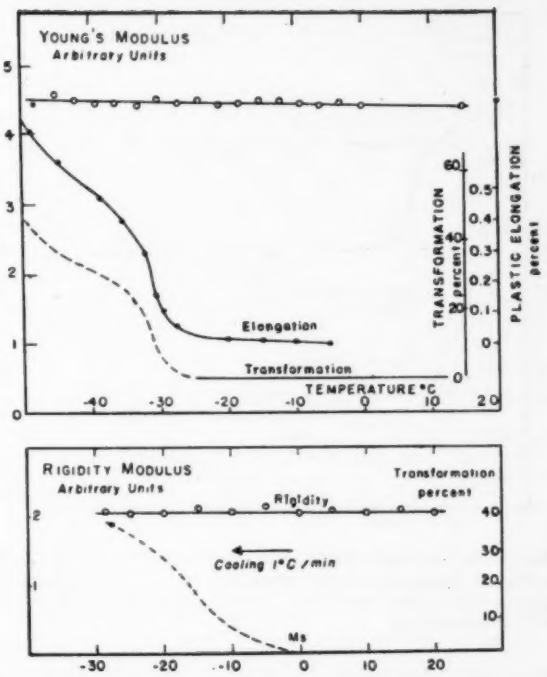


FIG. 3. Variation of static elastic moduli with temperature, in 1 mm diameter Fe-29 percent Ni wires intermittently stressed during cooling. Moduli are constant but plastic elongation under fixed load parallels amount of transformation.

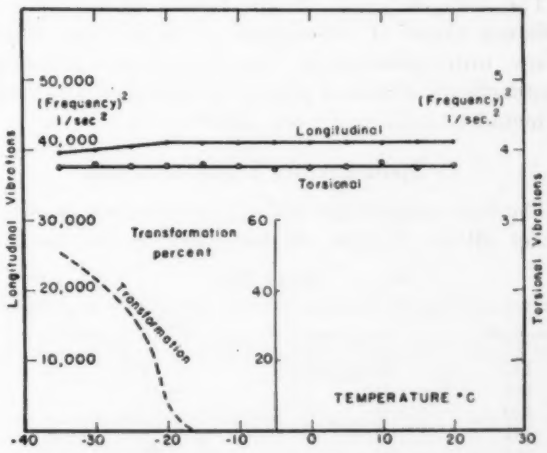


FIG. 4. Dependence of dynamic elastic moduli on temperature of 1 mm Fe-29 percent Ni wires continuously cooled. Longitudinal and torsional (natural frequencies) of a mass suspended on the wires are proportional to Young's modulus and shear modulus.

[14] A. W. McReynolds, J. App. Phys. 17, 823 (1946).

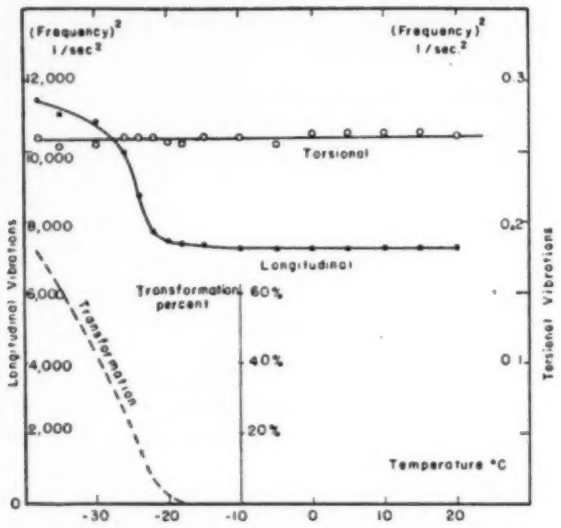


FIG. 5. Dependence of dynamic elastic moduli on temperature (see Fig. 4). Increase of longitudinal frequency as transformation begins indicates that the bent wire straightens as metal becomes soft in transforming.

system, the frequency of the oscillator input to the driving coil being adjusted to mechanical resonance as detected by a pick-up coil, amplifier and oscilloscope. Elastic moduli can then be determined from the relations

$$E = km\nu_l^2,$$
$$G = k'I\nu_t^2,$$

where E and G are Young's modulus and rigidity modulus, k and k' constants depending on dimensions of the specimens, m and I mass and moment of inertia and ν_l and ν_t longitudinal and torsional natural frequencies. As in other apparatus temperature was controlled by circulation of alcohol through a tube surrounding the specimen.

III. RESULTS

The data obtained are the results of a number of different types of measurements, which are divided below into spontaneous transformation by cooling, elastic effects, effects of plastic deformation, and results of metallographic and x-ray observation.

1. Spontaneous Transformation

The high temperature face-centered structure of iron-nickel alloys changes at low temperature to body centered. Fe-29 percent Ni alloys were used since that composition has a transition range slightly below room temperatures. To measure quantitatively the amount of transformation, electrical resistivity is the most convenient property, since it is found to be decreased by about 60 percent by the amount of transformation which occurs in quenching to −195°C. A typical curve of resistivity *vs.* temperature for a sample cooled at a uniform rate of 1°C/min. is shown in Fig. 2. The two straight lines represent resistivity of completely austenitic, and completely transformed, martensitic, structure and the connecting curve is resistivity of the mixture as transformation progresses. As discussed in a previous paper by the author,[14] the amount of transformation can be computed from the position of this transition curve between the two straight lines.

TABLE III.

Sample No.	Tensile stress	M_s temperature	
112	0 kg/mm²	−13°C	
113	0	−21	
129	0	−15	
131	0	−18	−17±3 Av
114	10	−20	
115	10	−15	
128	10	−20	
130	10	−16	−18±2 Av

The kinetics of the observed transformation are in accord with the characteristics of martensite formation in that there exists a fairly definite temperature (the M_s point), lying between 0° and −50° for the alloys used, at which transformation starts, and in that this temperature is not affected by changing the rate of cooling. Further, subsequent metallographic examination shows the (body-centered) product of transformation to be in the form of thin plates characteristic of the diffusionless martensite-type transformation. The needle-like regions in Fig. 16 are cross sections of these plates.

The reverse transformation, b.c.c. to f.c.c. occurs on reheating, but with considerable hysteresis, the transformation range lying from 400–470°. Resistivity changes during this reverse transformation are also shown in Fig. 2. The kinetics of the reverse transformation were found to be similar to those of the f.c.c. to b.c.c. reaction. That is, transformation continued only while heating, stopped if the temperature was held constant, then resumed when temperature next exceeded the previous maximum value.

In many cases of martensite formation a stabilization effect, as defined above, is caused by interrupting the cooling and holding the specimen at constant temperature either above the M_s point or during transformation. No such effect was found in the present experiments. On the contrary, the transformation exhibited a property not characteristic of martensite, in that once started it continued for some time after cooling was stopped and even when the temperature was raised a few degrees. Since this effect does not appear to bear directly on the main subject, its detailed discussion will be deferred to a later publication. This property of isothermal transformation does mean, however, that only the beginning point of transformation is a unique function of temperature, independent of cooling rate.

The size of the martensite plates which form are limited by the boundaries of the grains in which they appear. It seems possible also that restrictions of the grain boundaries might influence initiation of martensite plates making formation of a plate more difficult in

small grains, so that the transformation temperature might be expected to increase as boundary restrictions become less with larger grains. Very few studies of the effect of grain size on the M_s point have been made, however. Several samples were, therefore, treated to give widely different grain size, and the M_s points were compared. Annealing ½ hour in vacuum at 900°C gave grains of average diameter 0.030 mm. and 6 hours at 1300°C gave average diameter of 0.26 mm. To ensure that the differences observed were attributable only to grain size and not to changes of chemical composition during annealing, the same two samples were then recrystallized at 1300° and 900°, respectively, so that the relative sizes of the grains were reversed, and the M_s temperatures were again compared. The results are given in Table II. It will be seen that the M_s temperature varies about 30°C over the range of grain sizes covered.

2. Elastic Properties

A. Change of Elastic Moduli Near M_s Temperature

If decreasing temperature leads to thermodynamic instability as indicated in Fig. 1, a decrease in elastic moduli should result from flattening of the free energy curve as temperature approaches the M_s point. For this reason careful measurements of both Young's modulus and the rigidity modulus were made in the neighborhood of the M_s point, dynamically and statically. Young's modulus was measured under practically static conditions, by intermittently adding and removing increments of load on a specimen in the extensometer as cooling progressed, and rigidity modulus was measured by intermittent application of torque to a specimen in the galvanometer apparatus. Results of these tests are shown in Fig. 3. For dynamic measurement a weight assembly was hung on the bottom of the wire specimen and natural frequencies in longitudinal and torsional vibration measured continually as temperature decreased. Constancy of the elastic moduli, as computed from the natural frequencies, as the temperature decreases into the transformation range is shown in Figs. 4 and 5. Longitudinal vibrations of the weight assembly represent only elongation of the sample if it is perfectly straight, but if it is even slightly bent involves a combination of elongation and flexure. The natural frequency is consequently sensitive to slight bending of the 1-mm and 0.5-mm diameter wires in mounting, frequency being a maximum for a straight sample. Torsional frequencies, on the other hand, are almost independent of straightness and correspond to the true elastic modulus. Figure 6 shows that although torsional frequency remains constant with decreasing temperature, the progress of transformation allows straightening of the wire, increasing the longitudinal frequency and therefore the apparent modulus. In no cases, however, did the modulus decrease with decreasing temperature near or even within the transformation range.

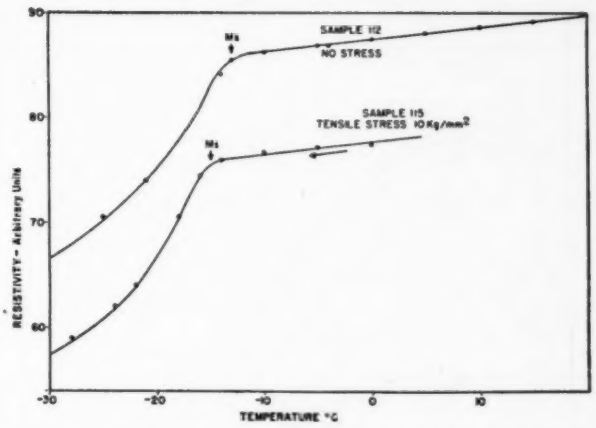


FIG. 6. Transformation of wires cooled with and without applied stress. Temperature (M_s) at which transformation begins is independent of stress within limits of error.

It is therefore, to be concluded that if thermodynamic instability occurs, it does so without appreciable change of the curvature of the free energy curve as shown in Fig. 1. This is in contradiction also to the observations of Scheil,[13] which indicated a decrease in statically measured Young's modulus near M_s.

On the other hand, the straightening effect mentioned shows that, as in other phase transformations, the structure becomes very soft during transformation; that is, it yields readily to external applied stress.

B. Effect of Elastic Stress on the M_s Point

It might be anticipated, since transformation is accompanied by plastic yielding, that the application of external stress would aid transformation and raise the transformation temperature slightly. To investigate this possibility a number of samples which had been given identical treatment were placed in the extensometer under fixed tensile stresses of 0.5, 2, 5, and 10 kg/mm²,

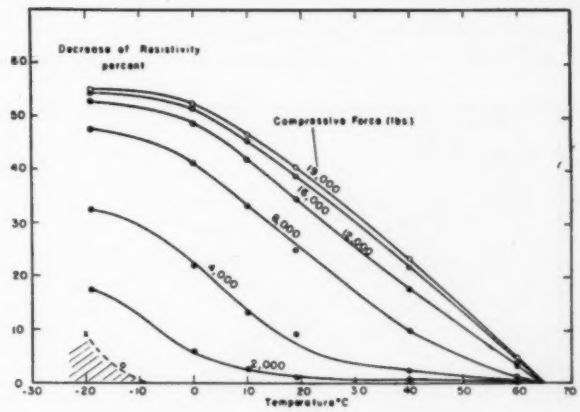


FIG. 7. Transformation by plastic deformation. Each vertical line of points represents successive compressions of 0.040 diameter×1.5.-in. Fe-29 Ni specimen between plates of hydraulic press; final thickness 0.015 in. Transformation approximately linear with decrease of electrical resistivity, 60 percent=100 percent transformation. Shaded region represents spontaneous transformation.

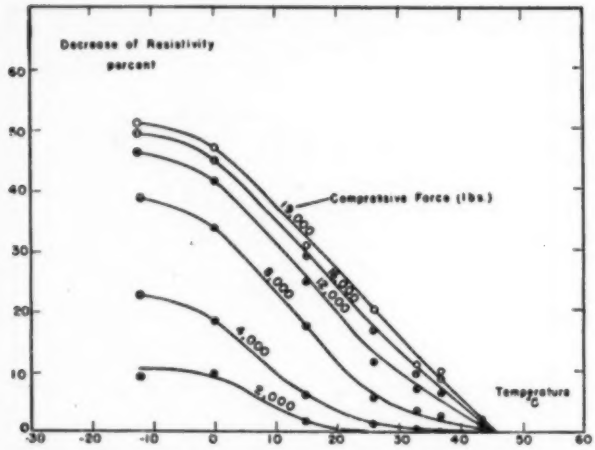


FIG. 8. Transformation by plastic deformation (same as Fig. 7), except average grain diameter (by linear count) 0.25 mm instead of 0.025 mm as in samples of Fig. 7.

and the M_s points determined by cooling at a uniform rate were compared. Stresses up to the elastic limit were found to have no effect. Table III shows a typical series of data on samples of large (0.25 mm) grain diameter.

3. Plastic Properties

A. Transformation by Plastic Deformation

A series of measurements was made of the amount of transformation which can be caused by plastic deformation. Wire samples 0.040-in. diameter × 1.7 in. were compressed in steps to 0.015-in. thickness between the compression blocks of the hydraulic press at various temperatures, controlled within 1°C. Procedure was to make successive compressions at 2000, 4000, 8000, 12,000, 16,000, and 19,000 lb. applied force, all at constant temperature, removing the load between compressions to make an electrical resistivity measurement. This method of plastic deformation allowed severe plastic flow without either fracture or significant change of the length or cross-sectional area. It was found, as shown in Fig. 7, that a large amount of transformation

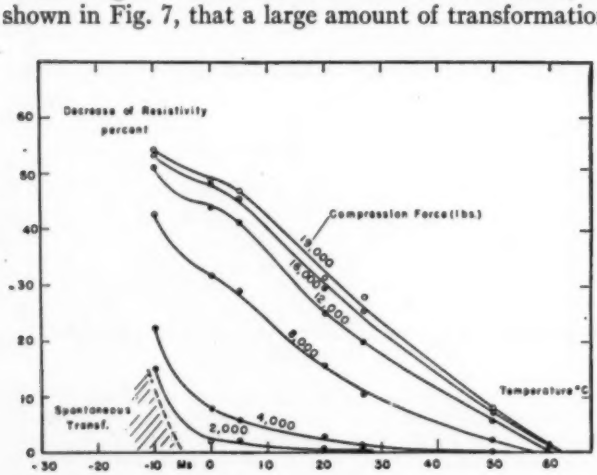


FIG. 9. Transformation by plastic deformation (same as Fig. 7) except samples originally square instead of round.

can be produced by plastic deformation and that the amount of transformation increases with the degree of deformation. Also, the amount of transformation for a certain degree of deformation is greatest near the M_s point and decreases uniformly with temperature, becoming zero at or above a rather definite temperature referred to hereafter as the "M_d point." Constancy of resistivity during compression at temperatures above this point is a convenient confirmation that the effects of dimensional changes and of cold-work on the untransformed matrix are negligibly small.

Figure 8, which is the result of a similar series of compressions on specimens of large grain size, shows no important difference from Fig. 7.

Since the method of compression used, cylindrical specimens between flat plates, obviously results in a very inhomogeneous strain, the small area in contact with the plates being severely strained before other parts of the specimen, some distortion of the relative positions of the curves in Fig. 7 was to be expected. To check on the magnitude of this effect a third series of compressions was made on specimens of the same wire which had been formed into square cross-sectional shape and re-annealed. The strain should in that case be nearly uniform throughout the specimen, at least for small compressions. Results of these tests are shown in Fig. 9 and again do not deviate significantly from Fig. 10 except that the 2000-lb. and 4000-lb. curves are somewhat lower, as expected.

B. Effect of Strain Rate on Transformation

In order to investigate the time relation between mechanical deformation and resultant transformation, continuous records of resistivity during compression were made. This was accomplished by inserting thin sheets of mica between specimen and compression plates for electrical insulation. Results showed that, as expected, the transformation is simultaneous with deformation and stops instantly when maximum pressure is reached except for a very slight additional transformation accompanying transient plastic creep of a few seconds' duration. These observations of course indicate simultaneity only within the resolving time, about 1 second, of the recording potentiometer. Time of deformation of the specimen in the hydraulic press was also about 1 second.

Compressions were also accomplished at very rapid rate by impact loading. The specimen was placed between the same flat blocks and a 20-lb. weight dropped from about 20-cm height on the top block. Time of compression under those conditions is estimated as about 0.005 second. These tests gave the surprising result that almost no transformation was produced, although the total deformation, if performed slowly, would have given about 50 percent transformation. Deformation at intermediate speed by passing the specimen wire between cylindrical rolls showed that compression in a time of 0.05 second caused slight transforma-

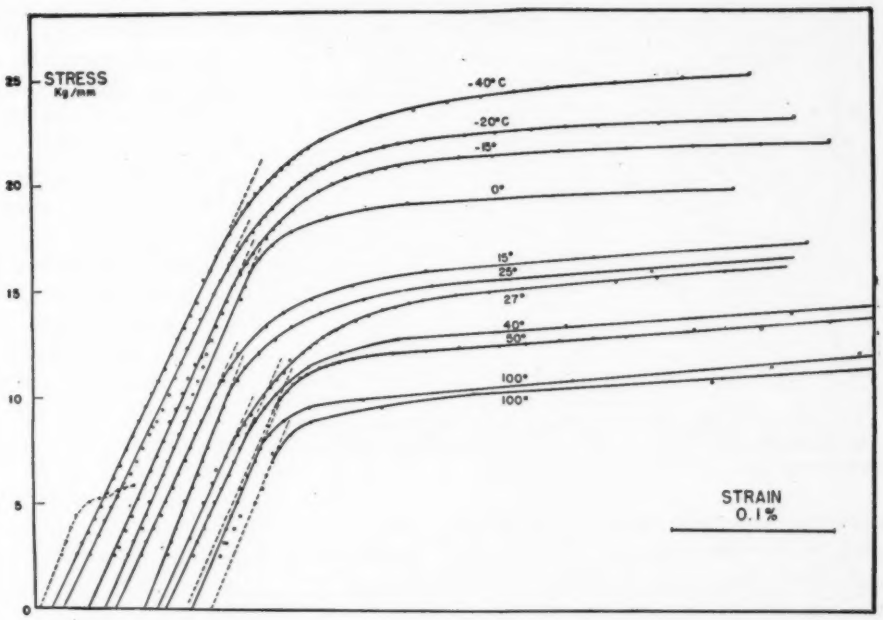


FIG. 10. Stress-strain curves for identical 0.040-in. diameter Fe-Ni samples strained at different temperatures in small tensile machine. Dotted curve is for specimen at −45°C just below M_s.

tion, and compression at 0.5 second had about the same effect as compression between plates in the hydraulic press. This effect of strain rate can probably be explained by the consideration that rapid deformation is adiabatic and quickly heats the specimen to a temperature above the M_d point, preventing further transformation. Furthermore, the heating would be expected to be localized at the slip bands where transformation presumably occurs. This explanation is substantiated quantitatively by both approximate calculations and experiments showing that the absorbed energy is sufficient to raise the temperature over 100°C. The alternative explanation, that there is a fundamental time delay in transformation, seems less likely.

C. Yield Stress Near M_s Point

If the mechanism of martensite formation is considered as a shear along some direction, then spontaneous transformation is an evidence of mechanical instability with respect to such a shear and the approach of instability with decreasing temperature should be represented by a series of free energy curves as represented in Fig. 1(A). Although results described above indicate that the elastic modulus remains unchanged, the "critical shearing stress" for transformation should certainly decrease with temperature. Thus the curves for yield by transformation and by slip should be as indicated in Fig. 1(B), and the intersection of the curves should represent the maximum temperature for martensite formation by deformation (designated here as the M_d point). Further, since transformation leads to macroscopically observable yielding of the specimen, the curve of yield stress *vs.* temperature should be as shown by the solid line in Fig. 1. The first part representing non-elastic yield by transformation and the second part non-elastic yield by ordinary plastic flow. In order to investigate the prediction, a number of stress-strain curves at various temperatures were determined by straining 3″ long wires of 0.040-in. diameter in the dial gauge extensometer. Results are plotted in Fig. 10. To have a consistent criterion of the initiation of yield, a deviation of 0.02 percent from the elastic curve was arbitrarily adopted as the yield point.

In Fig. 11 each point represents the yield point determined from one of the curves of Fig. 10. It will be seen that instead of decreasing according to the predicted curve as the temperature decreases below the M_d point the yield stress suddenly rises, the beginning of anomalous increase coinciding with the M_d point determined by compression tests on the same specimens. Thus the transformation which occurs by deformation not only

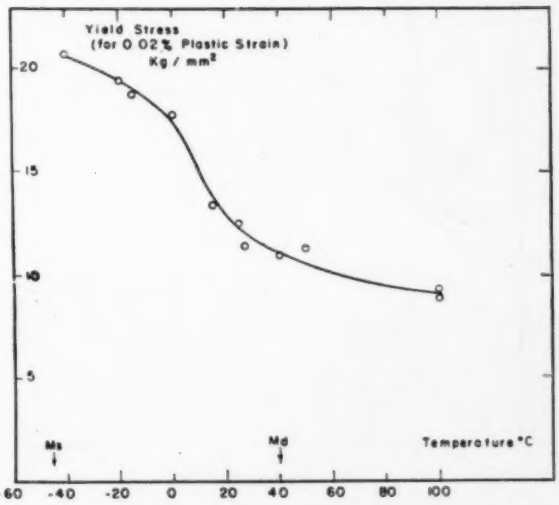


FIG. 11. Yield stress of austenitic Fe-Ni samples as a function of temperature. Each point derived from one of curves in Fig. 10. Note that sharp rise in the curve coincides with the M_d temperature.

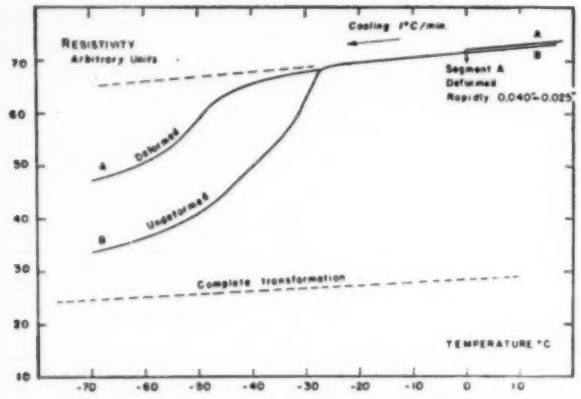


FIG. 12. Stabilizing effect of plastic deformation. Transformation begins 15–20°C lower in the compressed specimen than in the undeformed one.

does not supersede deformation by slip, but actually impedes it, strengthening the material. Simultaneous resistivity measurements during the strain showed that the amount of transformation resulting from these small strains is too small to be detected electrically (less than 2 percent).

D. Deformation of Body-Centered Cubic Phase

The similar kinetics of the γ- to α- and α- to γ-transformations suggests that deformation between M_d and the temperature for spontaneous α- to γ-transformation would convert the α-phase to γ. This point was investigated by quenching samples in liquid nitrogen then compressing them in the hydraulic press at temperatures up to 400°C. Resistivity was found to be unchanged by deformation, indicating no transformation.

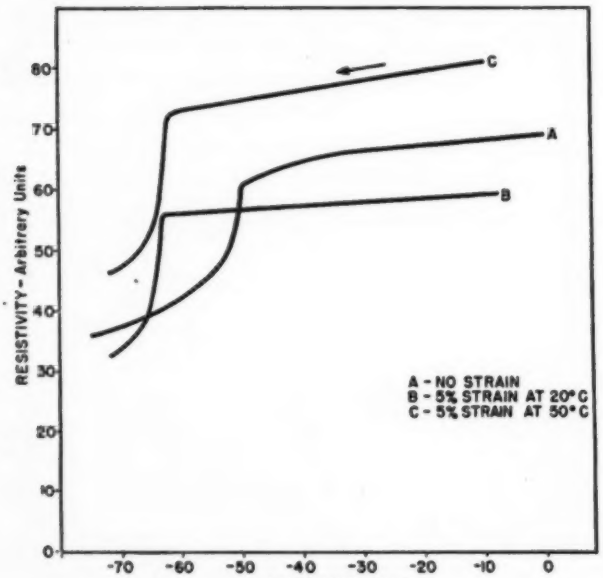


FIG. 13. Stabilizing effect of plastic deformation. Uniform 5 percent elongation depressed M_s temperature to the same extent, about 15°C, whether strain was performed at 20° or 50°C (below or above the M_d point).

4. Stabilization Effect

In addition to the effect of partly transforming the austenite above the M_s temperature, deformation also stabilized the austenite, preventing further spontaneous transformation during cooling. In Fig. 12 the resistivity *vs.* temperature curves of a previously deformed sample and a similarly treated but undeformed sample are shown. The principal effects of deformation are to depress the M_s temperature about 20°C and to decrease the sharpness of the beginning of transformation. It will be seen that depression of the M_s point is about the same for specimens deformed between the M_s and M_d points and those deformed above the M_d point. Thus stabilization is a consequence not of the partial transformation but of introduction of slip lines or other lattice imperfections.

A similar stabilization is observed in the reverse transformation from body-centered to face-centered cubic with increasing temperature. Measurements were made both by electrically heating specimens in vacuum and by immersion in a molten salt bath of gradually increasing temperature. Figure 14 shows the result of simultaneous measurements on three different specimens immersed in the same bath. The samples were adjacent segments of a wire converted to austenite by annealing $\frac{1}{2}$ hour at 900°C. Then segment A was transformed by quenching in liquid nitrogen, segment B quenched in liquid nitrogen then compressed at 20°C to 0.015 in. thickness, and segment C transformed by compression at -10°C (above M_s) to 0.015 in. Figure 14 shows that compression had the stabilizing effect of increasing the transformation temperature and spreading out the temperature range over which transformation occurs. Also, the stabilizing effect of the deformation is shown to be the same in the case where it induces the f.c.c. to b.c.c. transformation as in the case where transformation is completed by quenching prior to deformation.

5. X-Ray and Metallographic Examination

Although neither x-ray diffraction measurements nor metallographic examination are adapted to quantitative measurements of the degree of transformation in this case, they constitute a valuable confirmation of the qualitative features. X-ray patterns were made on samples which had been, respectively, annealed, quenched, and deformed by compression. The wire specimens were set up slightly off-center in a powder camera and patterns obtained by irradiation from a molybdenum target vacuum x-ray tube operated at 45 kev. Typical patterns are shown in Fig. 15. In the annealed specimen only face-centered lines are visible (15A); in a specimen quenched in liquid nitrogen the structure is almost entirely body-centered cubic (15C), and in a specimen strongly cold worked just above the M_s temperature, both sets of lines appear (15B). It was found that although transformation by severe deforma-

tion appeared to be almost complete according to resistivity measurements, the face-centered lines always persisted. In x-ray examination, about 0.005 in. of the surface was removed by electrolytic etching prior to exposure, since the structure of an outer layer about 0.001-in. thick differed from that of the interior, probably because of slight surface contamination during heat treatments.

Metallographic examination of samples partially transformed by cooling showed the usual structure of martensite needles (cross sections of plates) with random distribution but characteristic orientation, in the austenite grains. Needles were readily visible on the surface without etching (Fig. 16A). Scratches were made on the surface to determine by their displacement across needles whether a homogeneous shear occurred. Although distortion of the scratches was sometimes found, as in Fig. 16(A), the scratches usually remained relatively straight. Such examination yielded no clear evidence on the existence of shear, however, since the needles appear on close examination to be made up of clusters of much smaller ones, also shear might proceed preferentially perpendicular to the surface rather than parallel. In annealed samples a few needles were found in parallel, feather-like patterns extending from the twin boundaries (Fig. 16(B)) and in slightly strained samples many regions of martensite were visible, located preferentially on twin boundaries but also throughout the grains (Fig. 16(C)).

IV. DISCUSSION

The best recognized picture for the atomic motion in martensite transformation is a homogeneous shear of an entire plate-shaped region. The only proposal of a physical mechanism by which such shear occurs is that lattice stability decreases with decreasing temperature until the shear takes place spontaneously under the action of only internal stresses resulting from elastic distortions. According to the viewpoint of Scheil, stability at higher temperatures is so high that plastic yield by slip occurs before stresses reach the required level for transformation shear which results in a relative displacement of opposite faces of the plate. In this temperature region external stresses in the appropriate direction should become effective. In fact, since transformation can occur on several different planes, any applied stress should aid shear transformation on some sets of planes. These hypotheses lead to the following specific conclusions, which may be compared with experiment:

(a) Elastic moduli should be low just above M_s.

(b) The M_s temperature should increase slightly with increasing external applied stress.

(c) Yield stress should be very low just above the M_s temperature, increase to a maximum, then decrease slowly with increasing temperature.

(d) During spontaneous transformation the metal should yield to very small applied stresses.

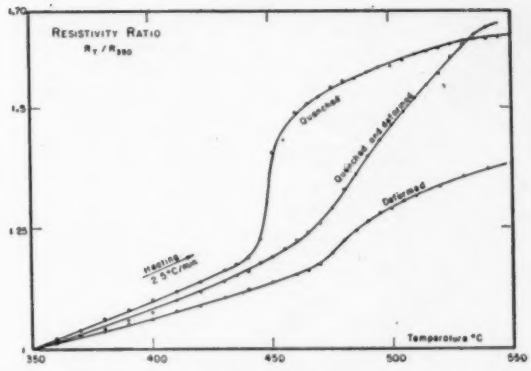


FIG. 14. Stabilizing effect of deformation on the b.c.c. to f.c.c. transformation. Temperature of reversion to f.c.c. is raised equally in specimens transformed by deformation, or deformed after transformation by cooling. Fe-29 Ni samples were compressed from 0.040 in. to 0.015 in.

(e) Transformation by deformation should occur between the M_s temperature and the maximum of the yield stress *vs.* temperature curve, but not at much higher temperatures.

The present observations of softness in the transformation range and transformation by plastic deformation in a limited region above M_s agree well with (d) and (e), but there is no evidence that (a), (b), and (c) are true. Predictions (b) and (c) can be reconciled with the results only by assuming that stability rises so rapidly with temperature that applied stresses are effective a few degrees above M_s. In that case, however, the deformation transformation up to 75°C above M_s is

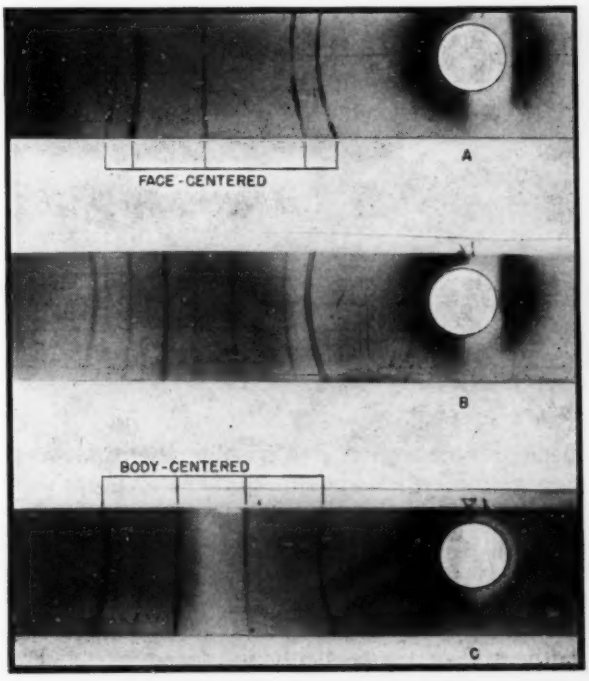
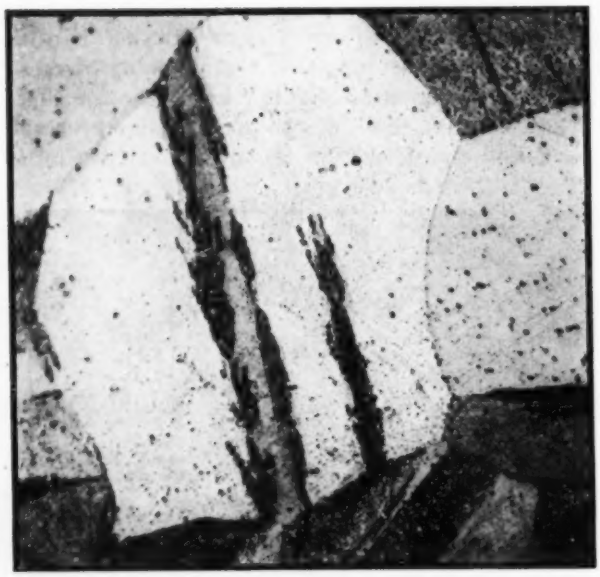


FIG. 15. X-ray diffraction patterns from Fe-Ni wires annealed ½ hour at 900°C. (A) no subsequent treatment; (B) compressed 0.040 in. to 0.015 in. at 20°C—partial transformation to b.c.c.; (C) Quenched to −195°C—almost complete transformation.

(A)

(C)



(B)

FIG. 16. Photo-micrographs of partially transformed Fe-Ni. (A) Quenched 5°C below M_s temperature, unetched-martensite visible in relief on surface. Distortion of fiducial scratches indicates some shearing motion. (B) Annealed at 900°C A few martensite needles appear as feather-like clusters on twin boundaries. (C) Annealed at 900°C and strained 10 percent. Dark etched martensite formed both at twin boundaries and throughout the grains.

difficult to account for. One possible alternative is to assume that there is a simultaneous homogeneous shear of two symmetrical regions such that no net displacement results. In that case, conditions (b) and (c) would not be expected to hold, but (d) and (e) are difficult to account for.

On the whole the present results do not support the hypotheses that transformation is a homogeneous shear or that it is effected by the action of shearing forces on a mechanically unstable lattice. It appears that slip and transformation are not alternative ways of yielding but are closely connected, since thermal transformation (by cooling) leads to yielding, and on the other hand deformation by stress results in isothermal transformation. It seems probable that the M_d point represents the temperature below which the body-centered phase is the more stable one. The beginning of spontaneous transformation at M_s is controlled by other factors; for example, the surface energy required in formation and spread of a martensite plate, and the availability of suitable nuclei.

If the lattice is violently deranged by slip, however, the displaced atoms fall into the more stable martensite structure. Furthermore, it is possible that these forcibly

deformed regions are large enough to serve as nuclei which spread into larger martensite plates. Appearance of the martensite as visible slabs, frequently arranged in a feather-like pattern extending out from the twin boundaries (Fig. 16) gives some support to this concept.

V. SUMMARY AND CONCLUSIONS

It is expected that the following characteristics of the martensite transformation in iron −29 percent nickel are also qualitatively applicable to martensite reactions in other materials:

1. No change in elastic moduli of austenite is associated with the decrease of temperature which leads to instability and transformation to martensite below the M_s point.

2. The M_s point, at which transformation begins, is not significantly affected by external applied stresses within the elastic limit for slip; that is, transformation above the M_s point cannot be induced by such stresses.

3. Plastic deformation of austenite at temperatures below or somewhat above the M_s point causes immediate transformation, the amount depending on the degree of strain and on temperature of straining. Transformation for about 60 percent compression varies linearly from 80–90 percent at M_s to zero at or above a temperature M_d, about 75°C above M_s.

4. The yield stress (for 0.02 percent plastic yield) begins to increase markedly with decreasing temperature at M_d and is at a maximum just above M_s. As the temperature is decreased from just above M_s to just below M_s, the yield stress changes discontinuously from a maximum to a very low value (Figs. 10 and 11).

5. Plastic deformation, at any temperature from below the M_s point to above the M_d point, has a stabilizing effect on the remaining austenite such that the beginning of subsequent transformation by cooling is depressed as much as 20°C and the temperature range over which transformation occurs is spread out.

6. An analogous stabilization in the reverse transformation from b.c.c. to f.c.c. is effected by plastic deformation of the martensite.

7. The M_s temperature varies appreciably with grain size, varying from −20°C for grains of 0.25 mm average diameter to −50°C for grains of 0.025 mm diameter.

8. The above properties give no support to the hypothesis that transformation occurs by a homogeneous lattice shear under the action of internal stresses when reduction of temperature renders the lattice mechanically unstable with respect to such a shear. They appear rather to indicate that the controlling factor is the presence of stable nuclei for martensite plates, the role of plastic deformation being to generate nuclei in deformed regions.

9. The M_d point apparently represents the temperature at which the b.c.c. phase becomes more stable than f.c.c. Transformation does not begin until temperature is further decreased to M_s because of surface energy effects (lack of nuclei of stable size) and possibly because the f.c.c. phase is still metastable.

ACKNOWLEDGMENTS

The author is indebted to Mr. John Gault for construction of much of the equipment and for making many of the measurements, to Messrs. H. Hirsch and G. Wilkerson for preparation of the alloys, and to Mr. O. Kammerer of Brookhaven National Laboratory for assistance in metallographic examination.

A High Temperature Precision X-Ray Camera

Some Measurements of the Thermal Coefficients of Expansion of Beryllium*

PAUL GORDON
Illinois Institute of Technology, Chicago, Illinois
(Received January 13, 1949)

A high temperature precision x-ray camera has been designed and constructed by modification of the familiar back-reflection symmetrical focusing type of camera. The camera, specimen, and furnace are used within a vacuum chamber to avoid oxidation. The apparatus in its present form has been employed at temperatures up to 1000°C, but higher temperatures could be easily attained by making a few minor changes in the furnace materials.

The lattice parameters of beryllium have been measured up to 1000°C. The derived coefficients of thermal expansion, both linear and bulk, are presented as a function of temperature. In addition, the data present strong evidence that the hexagonal close packed form of beryllium which exists at room temperature is stable up to at least 1000°C, contrary to several reports in the literature.

THE desirability of accurately measuring the lattice parameters of beryllium at elevated temperatures has led to the design and construction of a high temperature precision x-ray camera. Some data on the thermal expansion of beryllium have been obtained and are presented in this paper along with a description of the camera.

I. DESCRIPTION OF THE CAMERA

The camera is basically a modification of the familiar back-reflection symmetrical focusing design. The adaptation of this design to the needs of high temperature work is illustrated in the drawings of Fig. 1 and the photograph of Fig. 2. The rear segment of the camera which ordinarily holds the specimen was made removable. Thus, for room temperature measurements, this segment can be easily attached by means of set screws and the camera used in the normal way. For high temperature measurements, the rear segment is removed to make room for the furnace. In this case the specimen is held in its proper position on the circumference along which the slit and film lie by means of an adjustable supporting arm extending out from a post at the top center of the camera. To the specimen end of this arm is attached a vertical circular disk (stainless steel) having its rear surface machined to the same radius (2½ inches) as the camera. The disk is positioned so that its curved rear surface lies along the camera circle by pressing its flat front face against a spacer placed between the disk and the flat surfaces of the camera body exposed by the removal of the rear segment. As a result the specimen is in the proper position to satisfy the focusing conditions of the camera when it is held against the rear surface of the disk.

For reasons which will be given later, the beryllium specimen was made by pressing fine beryllium powder into one end of a longitudinal hole in a cylindrical beryllium holder, this end of the specimen holder being ground to fit the curved surface of the disk. The holder fits snugly into a hole in a cylindrical molybdenum rod (about 1-inch diameter, 1-inch long) with the face of the specimen flush with the properly curved face of the molybdenum cylinder. The latter is fastened to the stainless steel disk by two stainless steel screws and thus serves to hold the specimen in position over a hole drilled through the disk to allow passage of the x-ray beam.

The specimen is heated by a small resistance furnace consisting of a platinum winding on a 1-inch diameter Alundum core surrounded by a 1-inch layer of K-30 insulating brick, the whole being contained in a sheet metal cylinder. The transverse faces of the furnace are ¼-inch Transite. The furnace is, thus, a cylinder approximately 3½ inches in diameter and 1½-inch long with a tubular muffle of just the right size to enclose the molybdenum cylinder in which the specimen is held. The furnace is fixed in place by clamping it to the supporting arm described above.

In order to protect the specimen and the various heated parts from oxidation, the whole assembly consisting of the camera, specimen support, specimen, and furnace is placed in a vacuum chamber (see photographs in Figs. 3 and 4). The chamber is simply a length of 9-inch diameter copper tubing with a brass plate brazed to the front end (facing the x-ray beam) and a second, removable brass plate held in place at the rear by wing nuts. A rubber gasket coated with Celvacene high vacuum grease insures a vacuum tight seal between this plate and the copper chamber. The bottom of the chamber is fitted with a grooved adapter so that it may be clamped to the track of the Picker x-ray diffraction unit used as the x-ray source. The x-ray beam enters the vacuum chamber through a beryllium window (thickness=0.005 inch) in the front brass plate.

Using a Welsh Duo-Seal forepump and a metal oil diffusion pump (Distillation Products) pressures as low as 5×10^{-6} mm of Hg were obtainable with the furnace and specimen at room temperature. During

* This document is based on work performed on the Metallurgical Project, M.I.T. under Contract No. W-7405-eng-175 for the AEC.

high temperature runs, pressures of 10^{-5} mm of Hg were obtained after time was allowed for the degassing of the heated parts. In order to minimize the degassing period and its associated oxidation, CO_2 rather than air was introduced into the vacuum chamber each time it had to be brought to atmospheric pressure.

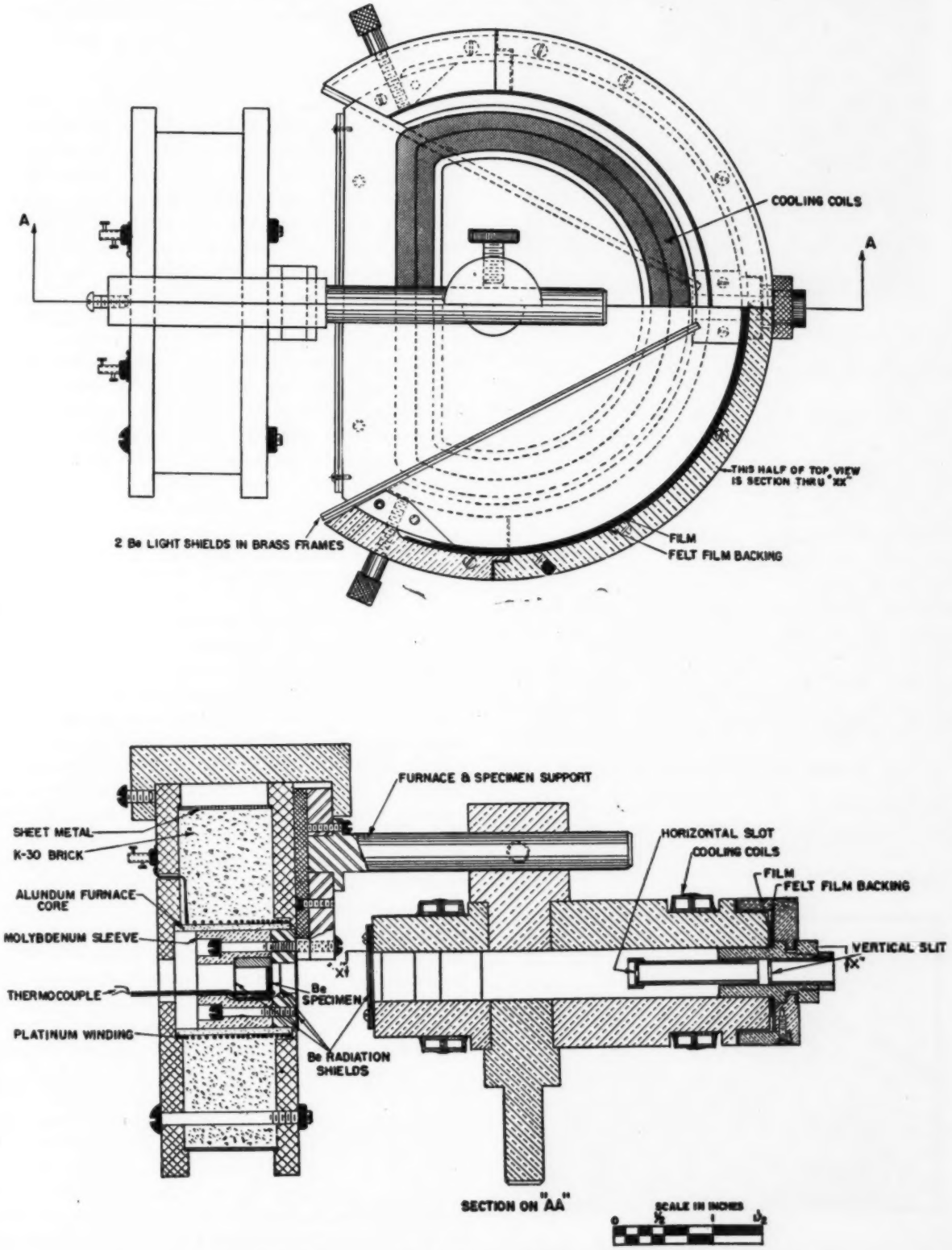


FIG. 1. Scale drawing of the high temperature precision x-ray camera.

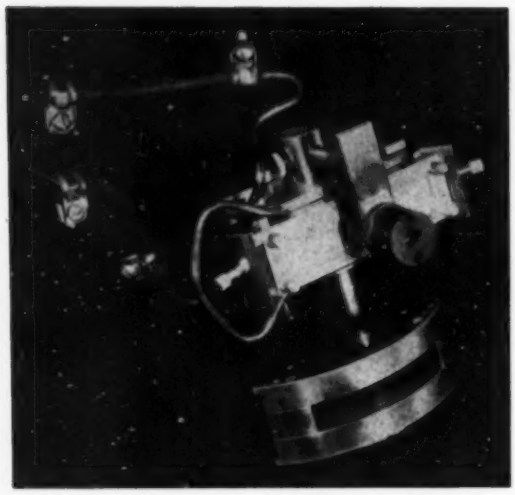


FIG. 2. View of the camera with high temperature attachments in place, but showing the removed rear segment which may be replaced for room temperature measurements.

The heat radiated by the furnace to the vacuum chamber walls and the camera is carried away by copper water-cooling coils. In the case of the chamber, these coils are soldered to the outside wall surfaces. The camera cooling coils are held tightly against the top and bottom camera surfaces by means of specially constructed clamps. The coils are connected to the water supply by rubber pressure tubing joined to short lengths of copper tubing, the latter being soldered through the vacuum chamber wall. In addition, a beryllium radiation shield (0.002-inch sheet Be) across the section of the camera facing the furnace serves to protect the photographic film in the camera against direct heat radiation from the furnace.

II. SPECIMEN PREPARATION

The preparation of beryllium specimens suitable for back-reflection x-ray measurements involves difficulties not encountered with heavier metals. With most metals a thin sprinkling of fine powder on a suitable backer,

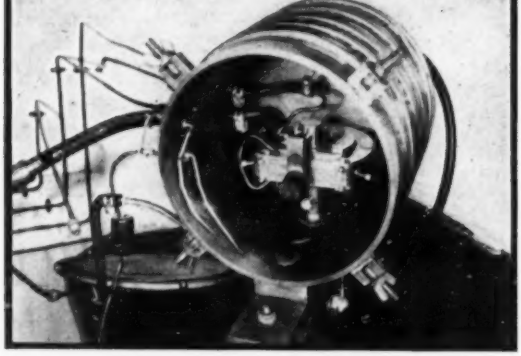


FIG. 3. View of the camera in place in the vacuum chamber, but with the furnace removed.

or a fine-grained solid piece with random grain orientation make satisfactory specimens. Because of the relatively high atomic numbers of such metals the x-rays penetrate only a very shallow surface layer of the specimen and an appreciable proportion of the x-ray energy is diffracted to the film. On the other hand, x-rays of the wave-lengths used in diffraction work easily penetrate as much as 5–10 mm of beryllium. This leads to several difficulties. First, it means that very little x-ray energy would be diffracted from a thin beryllium powder specimen and hence the exposure time would be prohibitively long. Second, although increasing the thickness of the specimen would shorten exposure time, the ultimate thickness to which one may go is limited by the necessity of meeting the focusing conditions of the camera. That is, for very thick specimens, the diffraction lines would tend to become broad, thus decreasing the accuracy of measurement, since the high penetrating power of the x-rays would result in diffraction not only from the surface, but from practically the full depth of the specimen. And third, the fact that a large portion of the x-radiation would be transmitted through even a very thick beryllium specimen means that anything behind the specimen would produce complicating diffraction lines and/or background intensity on the x-ray film. Thus, the space behind the specimen must be kept empty.

The most satisfactory beryllium specimen for use in a back-reflection powder camera would be one made from a fine-grained solid piece having random grain orientation. In this case the maximum amount of metal for a given specimen thickness could be placed in the path of the beam. However, the only relatively fine-grained beryllium available today is extruded material having a high degree of preferred orientation. The lack of randomness results in the loss of certain diffraction lines on the x-ray photogram. Consequently, the beryllium specimens for the present work were finally made by compressing very fine powder into the form of disks about 0.025-inch thick. Pressures of about 40–50,000 p.s.i. produced a density in the disks of approximately 60 to 70 percent that of solid Be. The powder was pressed into one end of a hole in a cylindrical beryllium holder as shown in the sketch of Fig. 5. A particle size of about 10 microns was used to obviate the necessity for rotating or translating the specimens. The cold-working stresses incurred during the pressing were relieved by annealing the specimen for 1 hour at 600°C in vacuum.

III. TEMPERATURE MEASUREMENT AND CONTROL

Measurement of the temperature of the specimen proved to be the most troublesome part of the procedure for making a high temperature picture. Because a thermocouple could not be attached directly to the pressed powder specimen, its temperature was measured by welding a Chromel-Alumel couple to the front face of the beryllium holder as close to the specimen as

possible (see Fig. 5). It was found to be very important to bring the thermocouple wires to the point of measurement *through* the furnace in order to eliminate heat losses from the hot junction by conduction through the thermocouple wires themselves. Failure to take this precaution was capable of producing errors in temperature determination of as much as 50°C, even with the very thin wire used. It was recognized that an appreciable temperature difference might exist between the position of the hot junction and the central part of the specimen on which the x-ray beam actually impinged. To minimize this difference two radiation shields of 0.002-inch Be sheet were placed across the furnace in front of the specimen and in contact with the heated stainless steel disk on the support arm. Under these conditions several calibration runs were made in an effort to measure the temperature difference between the hot junction and the center of the specimen. This was done by replacing the powder specimen with a solid, dummy specimen (in several cases molybdenum dummy specimens and molybdenum holders were used) spot-welding a second couple to the center of the front face of this dummy specimen, heating to various temperatures in vacuum and recording the difference in reading between the two couples. The difference was found to be less than 5°C for all temperatures up to 1000°C; an average correction as a function of temperature was obtained from the several calibration runs made and was applied to all subsequently measured temperatures.**

The thermocouples used were replaced for each new run, that is, each temperature recorded in the data (see Appendix) was registered with a fresh couple. This was done to eliminate the possible effects of contamination of the thermocouple wires, particularly by beryllium vapors. Test calibration of a once-used couple (used at high temperature) against a fresh one showed no significant difference in reading. All couples were made from the same batches of Chromel and Alumel wire, so that no appreciable *relative* error in temperature reading was introduced by the practice of using a new couple for each run.

Temperature control was obtained with the use of a voltage stabilizer and a Variac. Temperatures could usually be held to within 2–3°C in this way. Thus, the total temperature uncertainty for any run was seldom greater than about ±3°C (including that due to the uncertainty in the correction described above). It is believed, however, that the attaining of better temperature control would be advantageous in future work, since the largest part of the inaccuracies in the present measurement of parameter changes with temperature can be attributed to this uncertainty in the temperature of the specimen.

** This correction, of course, could be rigorously applied only to solid specimens; it was felt, however, that this represented as close an approach to the actual temperature of the powder specimen as could be obtained, and that the errors still unaccounted for were second-order effects.

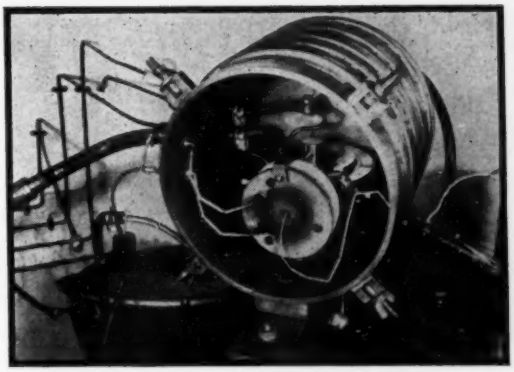


FIG. 4. View of the camera and furnace in place in the vacuum chamber.

IV. X-RAY TECHNIQUE

The back-reflection symmetrical focusing camera is an absolute instrument capable of giving lattice constants to an accuracy of 0.005 percent or better. No standard substance of known parameter need be used. For cubic crystals, the systematic errors due to camera radius error, absorption in the specimen, and film shrinkage may be eliminated by plotting the parameter calculated from individual diffraction lines against the function $\phi \tan(\phi/2)$ where $\phi=\pi-\theta/2$ and θ is the Bragg angle.[1] Extrapolation to $\phi \tan(\phi/2)=0$ gives the corrected value of the parameter. For non-cubic crystals, such as beryllium (HCP), the graphical extrapolation may not be used accurately, but an analytical method for extrapolating to eliminate the systematic errors has been developed by M. U. Cohen.[1]

Cohen finds, as indicated above, that symmetrical focusing cameras have systematic errors in $(\Delta d/d)$ (d=lattice spacing) that are proportional to $\phi \tan(\phi/2)$ in the range $\sin^2\theta>0.7$. Thus

$$\Delta d/d = D\phi \tan(\phi/2). \quad (1)$$

By squaring Bragg's law and taking logs we have

$$2 \log d = -\log \sin^2\theta + 2 \log(n\lambda/2). \quad (2)$$

Differentiating (2) and substituting the value of $(\Delta d/d)$ from (1) gives

$$\Delta \sin^2\theta = 2D[\phi \tan(\phi/2)]\sin^2\theta,$$

TABLE I. Wave-lengths of radiations used.

Radiation	Wave-length, angstrom units
Fe $\alpha 1$	1.93597
Fe $\alpha 2$	1.93991
Fe β	1.75654
Co $\alpha 1$	1.78890
Co $\alpha 2$	1.79279
Co β	1.62073

[1] M. U. Cohen, Rev. Sci. Inst. **6**, 68 (1935); **7**, 155 (1936); Zeits. f. Krist. **94**, 288, 306 (1936).

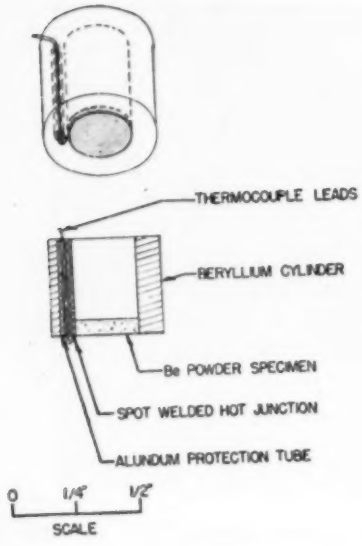


FIG. 5. Sketch of the specimen and specimen holder indicating method of attaching thermocouple.

and, since $\theta=\pi/2-\phi/2$, and $\sin^2\theta=\cos^2(\phi/2)$,

$$\Delta \sin^2\theta = D\phi \sin\phi.$$

Thus, the systematic errors lead to errors in the measured $\sin^2\theta$ values which are proportional to $\phi \sin\phi$. Adding this correction to the quadratic form for hexagonal crystals gives

$$\sin^2\theta = A(h^2+hk+k^2)+C(l^2)+D\phi \sin\phi$$

which may be written

$$\sin^2\theta = \alpha A+\gamma C+\delta D$$

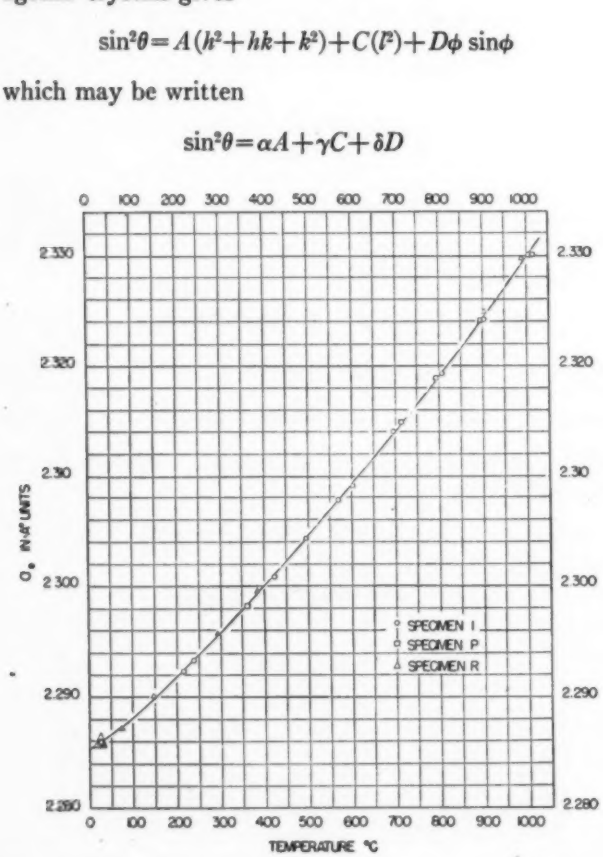


FIG. 6. Change in the a_0 parameter (edge of cell perpendicular to hexagonal axis) of beryllium as a function of temperature.

where $A=(\lambda^2/3a_0^2)$, $C=(\lambda^2/4c_0^2)$, $\alpha=h^2+hk+k^2$, $\gamma=l^2$, $\delta=5\phi \sin\phi$, and $D=$proportionality constant.

With a set of measured $\sin^2\theta$-values at hand for a given film, A and C may now be found by the method of least squares. The three normal equations are set up

$$\sum\alpha \sin^2\theta = A\sum\alpha^2+C\sum\alpha\gamma+D\sum\alpha\delta,$$
$$\sum\gamma \sin^2\theta = A\sum\alpha\gamma+C\sum\gamma^2+D\sum\gamma\delta,$$
$$\sum\delta \sin^2\theta = A\sum\alpha\delta+C\sum\gamma\delta+D\sum\delta^2,$$

the summation being carried out over all the equations of the separate diffraction lines. Solution of the simultaneous equations yields the corrected values of A and C, and thus, of the lattice constants a_0 and c_0. It will be found, however, that in most cases the solution of the simultaneous equations involves the loss of significant figures. The use of a simple mathematical device circumvents this difficulty and in addition makes it unnecessary to carry an inconveniently large number of significant figures in the calculations. The three normal equations are replaced by another set in the following way: Values of the constants A and C are calculated from the best available approximations of the lattice parameters a_0 and c_0. These values may be designated A_A and C_A, the subscript indicating "assumed." From A_A and C_A, assumed values of $(\sin^2\theta)_A$ can be obtained, neglecting the term $D\phi \sin\phi$ in the quadratic equation. Now, letting

$$\Delta A = A_A - A_E,$$
$$\Delta C = C_A - C_E,$$
$$\Delta D = D_A - D_E,$$
$$\Delta \sin^2\theta = (\sin^2\theta)_A - (\sin^2\theta)_E = v,$$

where the subscript (E) indicates the experimentally determined values, it will be seen that the new normal equations may be written

$$\sum\alpha v = \Delta A\sum\alpha^2+\Delta C\sum\alpha\gamma+\Delta D\sum\alpha\delta,$$
$$\sum\gamma v = \Delta A\sum\alpha\gamma+\Delta C\sum\gamma^2+\Delta D\sum\gamma\delta,$$
$$\sum\delta v = \Delta A\sum\alpha\delta+\Delta C\sum\gamma\delta+\Delta D\sum\delta^2.$$

Solution of these equations for ΔA and ΔC now yields values of A_E and C_E from the defining relationships above. Since ΔA and ΔC are small corrections to be applied to A_A and C_A to obtain the experimental values, only a few significant figures need be carried in the calculations.

The accuracy of the mathematical extrapolation described above is naturally the better the greater the number of diffraction lines with $\sin^2\theta>0.7$ appearing on the film. In the present work both Co and Fe radiations were employed for each photogram produced in order to obtain a sufficient number of diffraction lines. Making use of both the K_α- and K_β-wave-lengths (see Table I) from these two targets, 13 usable diffraction lines appeared on the film at specimen temperatures near room temperature, while at the higher temperature up to 16 lines were produced. Operating at 40 kv and 10 milliamp., the exposures required for satisfactory photograms were of the order of 1–2 hours for each target (2–4 hr. total for each film).

V. LATTICE DIMENSIONS OF BERYLLIUM AS A FUNCTION OF TEMPERATURE

The beryllium used for the present measurements was obtained in the form of −325-mesh powder from the Brush Beryllium Company. Powder particles having a size of about 10 microns or less were separated from the remainder in an infra-sizer. This very fine powder was then used to make the x-ray specimens. The quantities of impurities present in the separated powder are listed in Table II. The possible effect of these impurities on the lattice constants of beryllium is considered to be very small. It has been shown in previous work[2] that binary alloys of most of these elements (including the major impurities in the present beryllium, Al, Si, Mg) with beryllium containing 1 percent of the added element revealed only minute differences in lattice constants from those of pure beryllium even after quenching from 1000°C. Furthermore, much of the Mg, which was the largest metal impurity in the beryllium, was undoubtedly distilled off during the stress anneal in vacuum given each specimen before using. Only copper (of the listed elements) was found to be appreciably soluble in solid beryllium, but it is present in the beryllium powder in sufficiently small amounts to be unimportant. The powder assayed only about 97 percent beryllium by weight indicating, according to the analysts, the presence of about 2 percent by weight of oxygen (for which no reliable direct method of analysis is yet available). It is probable that a large part of this oxygen was picked up by the fine powder during ball milling, for experience has indicated that beryllium in finely divided form has a strong afinity for oxygen even at room temperature. However, here again previous work[2] suggests that oxygen, too, has little effect on the lattice dimensions of beryllium.

The order of magnitude of the total influence of the impurities (perhaps excluding the oxygen) on the beryllium unit cell dimensions may be indicated by the following figures: In all the previous measurements made in these laboratories on the purest beryllium available, the a_0 parameter has been found to be 2.2854 to 2.2856A at 25°C; the corresponding value for the powder, the analysis of which is given in Table II, is 2.2858_7A.

Measurements of the lattice dimensions of beryllium have been made at temperatures in the range of 25° to 1000°C*** on three separate specimens prepared from the Brush powder. The data obtained (listed in Table A of the Appendix) are plotted in the curves of Figs. 6–9. Figures 6 and 7 show the variation of the a_0 and c_0 dimensions of the unit cell as a function of temperature, Fig. 8 the resulting volume changes and Fig. 9 the trend in the axial ratio, c/a.

[2] P. Gordon, AEC Declassified Report No. MDDC 1370 (August, 1945).

*** The apparatus could be used at considerably higher temperatures by making a few minor changes in the materials used to build the furnace, particularly the Transite faces.

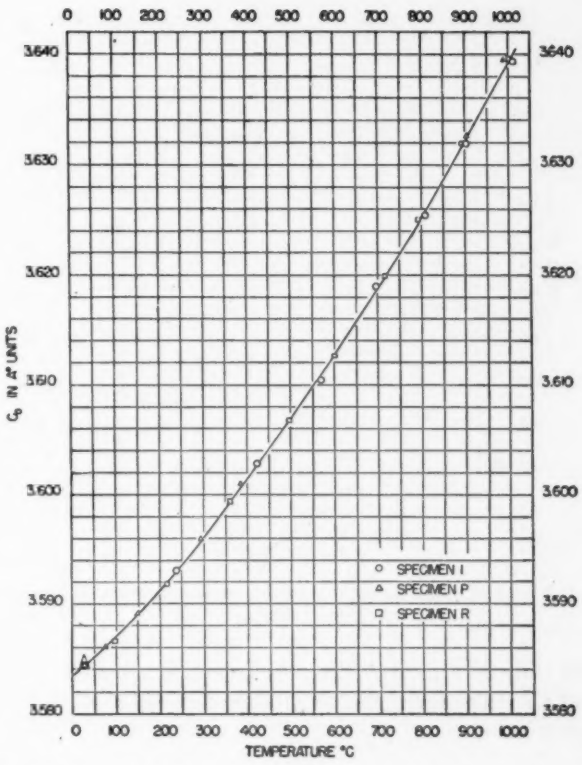


FIG. 7. Change in the c_0 parameter (length of the hexagonal axis) of beryllium as a function of temperature.

From these data it is possible to calculate the linear and volume coefficients of thermal expansion for beryllium. This has been done and the results plotted in the curves of Figs. 10 and 11 as a function of temperature. In Fig. 10 are represented the *true* coefficients of expansion calculated from the slopes of the a_0, c_0 and volume—*versus* temperature—curves, respectively, at any given temperature. Thus, the values given on the ordinate of the graph in Fig. 10 are values of $1/x(dx/dT)$, where x is either the a_0 or c_0 dimension or the unit cell volume at the temperature, T. The slopes, dx/dT, were found by drawing the best curves through the experimental data (Figs. 6–8), picking values of, say, the a_0 parameter from the proper curve at points 50° above and 50° below the temperature at which the slope was desired, and dividing the difference between the two values thus obtained by the temperature difference, 100°C. The volume coefficients were calculated

TABLE II. Major impurities in the beryllium powder.

Element	Wt., %	Atomic %
Fe	0.094	0.016
Al	0.280	0.104
Si	0.141	0.047
Mg	0.580	0.23
C	0.06	0.05
Ca	0.07	0.02
Cu	<0.01	<0.002
Mn	0.04	0.007
Ni	0.01	0.002

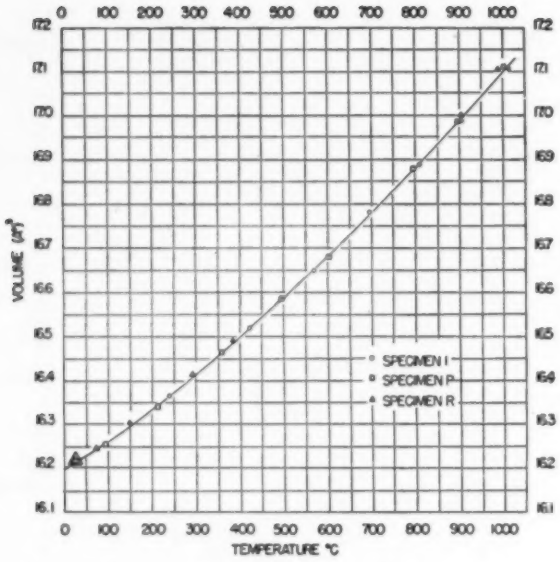


FIG. 8. Change in volume of the beryllium unit cell as a function of temperature.

not only in this way directly from the data in Fig. 8 but also from the linear coefficients derived as above and the relationship

$$\alpha_v = 2\alpha_\perp + \alpha_{||},$$

where α_v = volume coefficient of expansion, $\alpha_\perp$ = linear coefficient of expansion $\perp$ to hexagonal axis, and $\alpha_{||}$ = linear coefficient of expansion $||$ to hexagonal axis. Consequently, two sets of points, corresponding to the two methods of calculation, are shown on the volume curve in Fig. 10. The good check of these two sets of values is an indication that the three basic curves in Figs. 6–8 have been consistently drawn through the experimental points.

These coefficient curves are, perhaps, of the greater scientific interest, but from the practical point of view the curves shown in Fig. 11 should be more useful. Here are plotted the more generally used coefficients of thermal expansion $\alpha_\perp$, $\alpha_{||}$ and α_v, as defined by the equations

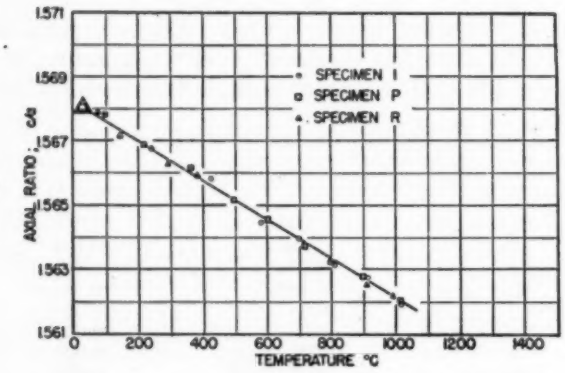


FIG. 9. Change in the axial ratio c_0/a_0, of beryllium as a function of temperature.

$(\text{length})_T = (\text{length})_{25}(1+\alpha_\perp T)$ $\perp$ to the hexagonal axis,
$(\text{length})_T = (\text{length})_{25}(1+\alpha_{||} T)$ $||$ to the hexagonal axis,
$(\text{volume})_T = (\text{volume})_{25}(1+\alpha_v T)$.

Although the volume coefficients of expansion are applicable to beryllium specimens of any type, it should be remembered that the linear coefficients can be rigorously applied only to single crystals where the direction of the expansion with respect to the hexagonal axis is known. However, the values read from the curves may be used as very close approximations for extruded round rods of beryllium where the extrusion reduction in area was greater than about 6 or 8 to 1. It has been demonstrated[2] that such material has a high degree of preferred orientation, the basal (0001) planes lying within about 6° or 7° of parallelism with the rod axis, but being randomly rotated about this axis. Thus, the longitudinal coefficient of thermal expansion for extruded rods is given very closely by the coefficient perpendicular to the hexagonal axis of the unit cell; the coefficient of radial thermal expansion for the rods is the arithmetical average of those $\perp$ and $||$ to the hexagonal axis; that is $(\alpha_\perp+\alpha_{||})/2$. For a polygrained specimen exhibiting no preferred orientation, the coefficient of linear thermal expansion in any direction may be taken as equal to $(2\alpha_\perp+\alpha_{||})/3$.

The maximum relative error in the individual lattice constants plotted in Figs. 6 to 9 is believed to be within ± 0.01 percent, the major part of this error being the result of temperature uncertainty which ranged from $\pm\frac{1}{2}$°C to ± 5°C (see Appendix, Table A). These small errors in the parameters, however, lead to comparatively large errors in true coefficients of expansion which are based on the slope of the parameter-temperature curves. It is estimated that the calculated true coefficients are accurate to no better than ± 5 percent at the extremities of the temperature range investigated, but are considerably less uncertain at the intermediate temperatures. The mean coefficients, on the other hand, are relatively inaccurate only at the lowest temperatures where the parameter change from 25°C to the temperature in question is small. The accuracy is probably about ± 5 percent at these low temperatures, approaching ± 0.2 to 0.3 percent as the temperature nears 1000°C.

The high temperature x-ray measurements have served not only to determine the coefficients of thermal expansion of beryllium, but have also supplied evidence on the question of whether or not beryllium undergoes an allotropic transformation at some elevated temperature. Such a transformation in beryllium has been reported in the literature on several occasions. Lewis[3] found indications of allotropic transformations at about −45°C and 450°C by thermal e.m.f. and electrical resistivity measurements. Noyce and Daane[4] reported a

[3] E. J. Lewis, Phys. Rev. 34, 1575 (1929).
[4] W. K. Noyce and A. H. Daane, AEC Report No. CT-2404 (March 15, 1945).

transformation at 730°C on the basis of thermal and dilatometric effects, both effects being very small. Attempts by Dr. A. R. Kaufmann and the present author to check the results of Noyce and Daane by careful thermal analysis gave no evidence of the alleged transformation,[5] however. In addition, Jaeger *et al.*[6,7] reported a transformation in Be based on x-ray and specific heat measurements. They gave the structure of the high temperature phase as hexagonal close packed with a relatively large unit cell ($a_0=7.1$A, $c_0=10.8$A, 60 atoms per unit cell). Their x-ray measurements were made at room temperature, but they indicated that the high temperature phase was most successfully produced by heating to 630°C.

The present high temperature x-ray measurements give strong evidence against the occurrence of a phase transformation at any temperature between 25°C and 1000°C. In Fig. 12 are reproduced several x-ray photograms taken at temperatures within this range. (The extraneous spotty lines in these pictures are due to the beryllium radiation shields between the specimen and the x-ray source. Since these spots were, in general, not superimposed on the lines from the specimen, they led to no measurement difficulties.) In each case the specimen was generally at the indicated temperature for about 2 to 3 hours before the exposure was started (to allow stabilization of the temperature) and then from 2 to 4 hours more during the exposure. It seems highly probable that, in these periods of time at the higher temperatures the formation of a stable high temperature phase would be sufficiently far advanced to be detectable in the diffraction pattern. If such a transformation had in fact taken place, the diffraction patterns would be expected to reveal a new set of lines or a sudden large discontinuous shift in the lines as the temperature was raised. Actually, as can be seen in Fig. 12, the lines of the room temperature hexagonal close packed beryllium phase persisted with essentially undiminished intensity at all temperatures up to 1000°C. They underwent a steady shift in position toward lower angles as the temperature was raised, consistent with a gradual increase in lattice parameter, and a few new lines of higher indices appeared on the high angle portion of the film as the increasing temperature brought the $\sin^2\theta$-values for these lines below unity. No sudden shift in line position was noted at any temperature, nor was there any evidence of a new set of lines corresponding to a high temperature phase. At 1000°C, however, there appeared on the film for all three specimens a faint alpha-doublet and a very faint beta-line (see Fig. 12) which were not present at any of the lower temperatures. It was immediately suspected that these were BeO lines resulting from the formation of an oxide film on the surface of the beryllium specimens. That this was actually the case was conclusively proven by making a diffraction pattern of pure BeO at 1000°C using Fe and Co radiation. The position of the strongest alpha-doublet and the strongest beta-line appearing on the BeO pattern thus obtained checked exactly with those of the corresponding lines on the 1000°C beryllium film.

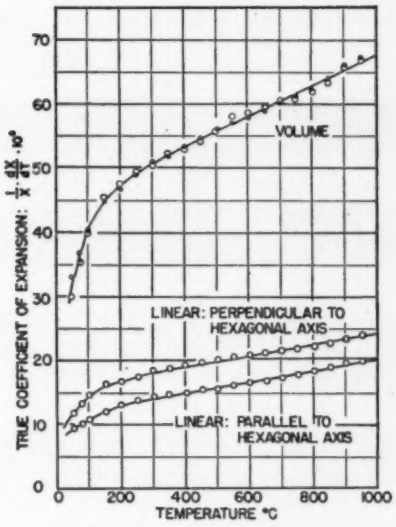


FIG. 10. True coefficients of thermal expansion of beryllium.

As has been indicated, the metal powder used in the above experiments assayed only about 97 percent beryllium. There remained, thus, a small probability that the presence of the impurities (largely oxygen) may have inhibited the formation of a high temperature phase which would normally be present in purer beryllium. Fortunately, some much higher assay beryllium powder recently became available. This powder was ground from vacuum cast beryllium by special

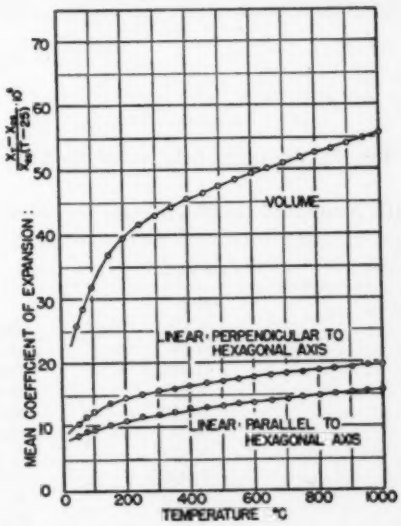


FIG. 11. Mean coefficients of thermal expansion of beryllium.

[5] P. Gordon and A. R. Kaufmann, AEC Report No. CT-3379 (December 11, 1945).
[6] F. M. Jaeger and J. E. Zanstra, Proc. Amst. Akad. Sci. **36**, 636 (1933).
[7] F. M. Jaeger and E. Rosenbloom, Proc. Amst. Akad. Sci. **35**, 1032, 1055; **57**, 67 (1934); Rec. Trav. Chim. Pays-Bas **53**, 451 (1934).

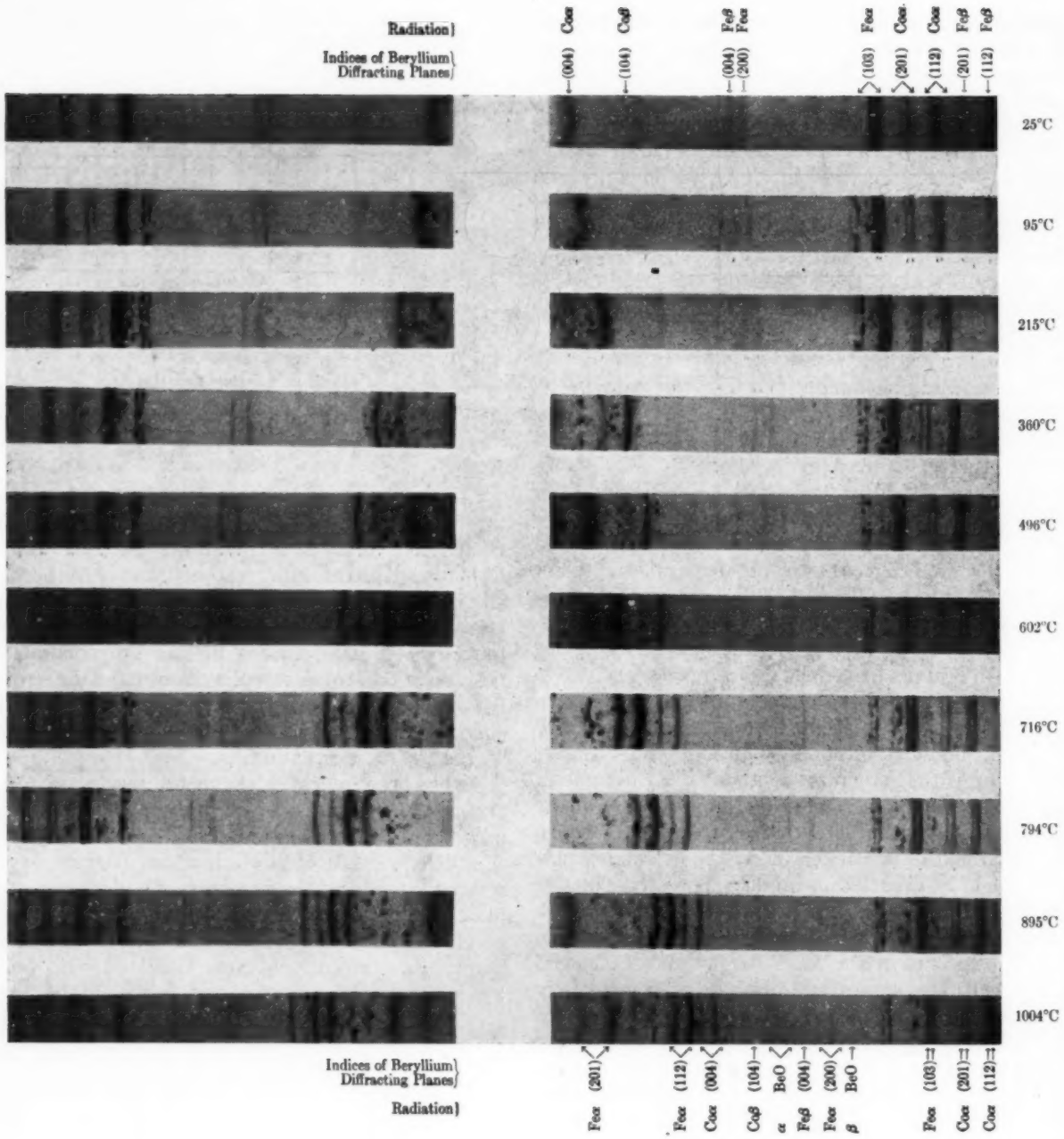


FIG. 12. Variation of the back reflection powder diffraction pattern of beryllium as a function of temperature.

TABLE III. Analysis of the 99.4% assay beryllium powder.

Element	Wt., %	Atomic, %
Be assay	99.4	
C	0.12	0.091
Al	0.17	0.057
Ca	0.014	0.0032
Cr	0.014	0.0024
Cu	0.010	0.0014
Fe	0.20	0.033
Mg	0.083	0.0031
Mn	0.020	0.0033
Ni	0.010	0.0016
Si	0.11	0.036

techniques devised at Brush Beryllium Company with the specific purpose of preventing the contamination of the powder by oxygen. The quantities of impurities present in the powder are indicated in Table III. Since the metal assayed 99.4 percent Be and the total impurity content (largely Al, Fe, Si and C) approximated 0.7 percent, it may be seen that the powder was essentially free of oxygen.

An x-ray specimen was prepared from this powder as received (−325 mesh) taking care to isolate the powder from contact with air as much as possible until it had been pressed into pellet form by the techniques

described earlier. After a 600°C stress relief anneal in vacuum, x-ray photograms were made at 650°C and 900°C. In neither case was there any indication of the formation of a high temperature phase.

It is considered that these experiments provide persuasive evidence that the hexagonal close-packed form of beryllium encountered at room temperature is stable up to at least 1000°C.

ACKNOWLEDGMENTS

The writer wishes to express his sincere thanks to Dr. A. R. Kaufmann for his continued helpful interest and to Herbert Sawyer whose assistance in the laboratory was invaluable.

APPENDIX

TABLE A. Measured lattice constants of beryllium.

Specimen	Temperature °C Av.	Uncertainty	a_0 in A	c_0 in A	c_0/a_0	Unit cell volume $(A)^3$
R	24	± ½	2.2858_7	3.5842_8	1.5680_2	16.219_4
P	25	½	2.2858_7	3.5843_1	1.5680_3	16.219_6
I	26½	½	2.2858_8	3.5843_6	1.5680_4	16.219_9
R	74	½	2.2871_4	3.5860_3	1.5679_1	16.245_4
P	95	1½	2.2875_6	3.5865_9	1.5678_7	16.253_9
R	147	2	2.2900_9	3.5890_4	1.5672_1	16.301_0
P	215	1	2.2923_2	3.5918_2	1.5668_9	16.345_6
I	239	2½	2.2932_4	3.5930_0	1.5667_8	16.364_0
R	293	3	2.2958_0	3.5959_9	1.5663_4	16.414_0
P	360	1½	2.2982_4	3.5994_2	1.5661_6	16.464_7
R	385	2	2.2995_2	3.6009_9	1.5659_7	16.490_3
I	423	1½	2.3008_8	3.6027_9	1.5658_3	16.518_1
P	496	1	2.3043_8	3.6067_8	1.5651_9	16.586_6
I	567½	2½	2.3078_5	3.6105_8	1.5644_8	16.651_6
P	601½	2½	2.3091_0	3.6127_1	1.5645_5	16.682_2
I	695	3½	2.3140_3	3.6190_6	1.5639_3	16.783_5
P	716	3½	2.3148_9	3.6199_3	1.5637_6	16.799_2
P	793½	3	2.3189_3	3.6251_4	1.5632_8	16.882_3
I	807½	2½	2.3193_9	3.6255_5	1.5631_5	16.890_8
P	895	3½	2.3241_0	3.6220_8	1.5627_9	16.989_9
I	901½	3	2.3241_9	3.6220_4	1.5627_3	16.991_2
R	904½	3½	2.3249_1	3.6227_4	1.5625_4	17.005_1
R	989½	4½	2.3298_4	3.6296_9	1.5622_2	17.109_9
P	1004	5	2.3300_8	3.6297_5	1.5620_8	17.113_7
I	1013	3½	2.3300_7	3.6294_6	1.5619_6	17.112_2

TABLE B. True coefficients of thermal expansion: $\alpha=1/x(dx/dT)\times10^6$ as calculated from Figs. 6–8.

Temperature °C	$\alpha_\perp$*	$\alpha_\parallel$**	α_v*** From Fig. 8	$2\alpha_\perp+\alpha_\parallel$
50	11.7	9.4	29.8	32.9
75	13.3	10.0	35.3	36.5
100	14.7	10.8	39.9	40.2
150	16.2	12.1	45.1	44.6
200	16.7	13.1	47.4	46.5
250	17.4	13.9	49.2	48.8
300	18.2	14.4	50.3	50.8
350	18.5	14.5	52.0	51.6
400	19.1	14.9	53.0	53.1
450	19.6	15.2	54.1	54.4
500	20.1	15.6	55.8	55.9
550	20.4	16.2	58.0	57.0
600	20.6	16.5	58.5	57.7
650	21.0	16.9	59.5	58.9
700	21.4	17.4	60.6	60.1
750	21.6	17.8	61.0	61.1
800	22.0	18.3	61.9	62.3
850	22.4	19.0	63.2	63.9
900	23.2	19.6	65.7	65.9
950	23.6	19.8	67.0	67.1

* $\alpha_\perp$ = Coefficient perpendicular to hexagonal axis.
** $\alpha_\parallel$ = Coefficient parallel to hexagonal axis.
*** α_v = Volume coefficient.

TABLE C. Mean coefficients of thermal expansion: $\alpha=[x_T-x_{25}/x_{25}(T-25)]\times10^6$ as calculated from Figs. 6–8.

Temperature °C	$\alpha_\perp$*	$\alpha_\parallel$**	α_v***
50	10.7	8.7	26.0
75	11.7	9.4	28.5
100	12.6	9.6	32.0
150	13.9	10.4	37.0
200	14.7	11.0	39.5
250	15.1	11.6	41.6
300	15.7	12.1	43.0
350	16.1	12.5	44.4
400	16.5	12.8	45.5
450	16.8	13.1	46.5
500	17.2	13.3	47.4
550	17.5	13.6	48.6
600	17.8	13.8	49.5
650	18.0	14.1	50.2
700	18.3	14.3	51.2
750	18.5	14.5	52.0
800	18.7	14.8	52.7
850	19.0	15.0	53.4
900	19.2	15.3	54.2
950	19.5	15.5	55.0
1000	19.7	15.8	55.8

* $\alpha_\perp$ = Coefficient perpendicular to hexagonal axis.
** $\alpha_\parallel$ = Coefficient parallel to hexagonal axis.
*** α_v = Volume coefficient.

Microwave Q Measurements in the Presence of Series Losses

L. MALTER* AND G. R. BREWER**
Naval Research Laboratory, Washington, D. C.
(Received January 17, 1949)

At very high frequencies, the effect of losses in coupling devices makes itself felt in the determinations of the various Q's and of the circuit efficiency of resonant systems. It is shown that the neglect of these losses (which appear in equivalent circuits in the form of series resistance), can lead to appreciable error. Formulas are derived and curves are plotted which permit an unambiguous determination of the Q's and of the circuit efficiency when the following items are known:

1. the standing wave ratio at resonance,
2. the standing wave ratio far from resonance, and
3. knowledge as to whether the resonant system under study is "under" or "overmatched" to the external load.

INTRODUCTION

IN the study of microwave circuits involving resonant elements, it is often of considerable value to determine the various Q values associated with the elements.

Microwave systems are distributed over dimensions comparable with the wave-length. However, if one restricts consideration to frequency regions in the neighborhood of resonances of the system, the behavior of microwave systems is similar to that of circuits with properly chosen lumped constants. In some cases of resonant microwave systems, the behavior in the neighborhood of resonance is similar to and can in fact be represented by that of a parallel resonant circuit consisting of an inductance, capacitance and conductance in parallel; whereas in other cases, the behavior is like that of a series resonant circuit. The circuits or systems exhibiting either series or parallel resonant properties can be transformed one into the other by means of a suitable coupling device such as a quarter-wave transformer.

The method of determining the impedances and Q's of such systems was developed extensively during the war.[1] In actual practice standing wave measurements made in a transmission line or wave guide terminated by the unknown impedance, permit the unknown impedance and Q's to be determined. It has been customary to assume that (for the parallel resonant case) the equivalent circuit of Fig. 1 is generally applicable. In actual practice, particularly with increasing frequency, such is not always the case. The departure from the ideal case arises largely due to resistive losses in the coupling system, or in the transmission line or waveguide in which the measurements are made. It is the purpose of this paper to discuss the manner in which the effect of these losses may be taken into account in the determination of the Q's and of the circuit efficiency.

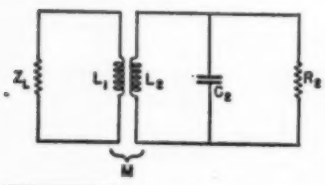


FIG. 1. Equivalent circuit of a resonant microwave cavity coupled to an external load.

THE CASE OF SERIES LOSS

Let it be assumed that a system represented by the circuit of Fig. 2 is under consideration. Such a system might be a cavity or magnetron coupled to an external transmission line by means of a loop. The resistance r_s represents the major series losses which are those present in the loop and the line. Other series losses, such as those in the secondary, are in general negligible or may be considered as having been transferred to the primary where they are included in r_s. The circuit of Fig. 2 can be transformed into that of Fig. 3.[2] The term $L_0=((L_1L_2/M^2)-1)L_2$ when multiplied by $j\omega$ is the so-called leakage reactance. Let it be assumed that through the frequency range of interest, the leakage reactance is substantially a constant. In that case, Fig. 3 can, in turn, be transformed into Fig. 4. Z_L is the load impedance and is unchanged as we go from Fig. 2 to Fig. 3 to Fig. 4. The exact determination of L, C, and R from L_1, L_2, M, C_2 and R_2 is complex. Fortunately, it need not be gone into, as the actual quantities of interest are L, C, and R, and the method of measure-

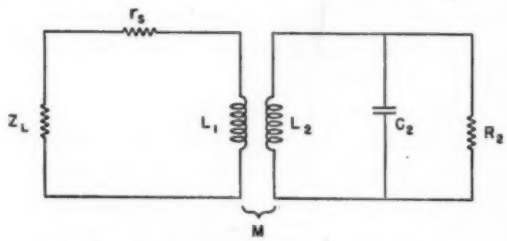


FIG. 2. Parallel resonant circuit coupled into load through a lossy transformer.

* Now at RCA Laboratories Division, Radio Corporation of America, Princeton, New Jersey.
** Now at Engineering Research Institute, University of Michigan, Ann Arbor, Michigan.
[1] J. L. Lawson, *Measurement of Impedance with the Standing Wave Detector* (MIT Report 64-3, May 18, 1942).
[2] E. A. Guillemin, *Communication Networks* (John Wiley & Sons, Inc., New York, 1935), Vol. 2, p. 156.

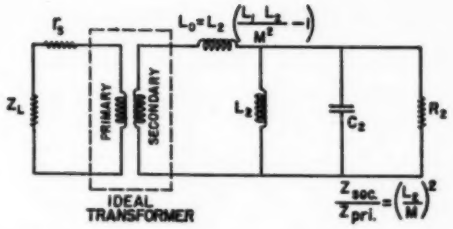


FIG. 3. Equivalent circuit of a resonant cavity or transmission line coupled to a load through a lossy transformer.

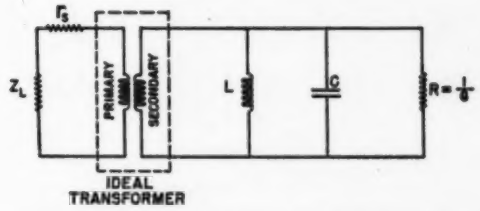


FIG. 4. Equivalent circuit wherein the leakage reactance is substantially uniform over the frequency range of interest.

ment to be described permits either a direct determination of L, C, and R or a determination of their values multiplied by $(L_2/M)^2$ or its reciprocal.

The load impedance Z_L can be transformed through the ideal transformer of Fig. 4 to the secondary side. The equivalent circuit then takes the form shown in Fig. 5. The circuit of Fig. 5 certainly does not represent all possible microwave systems.

However, it is applicable to many more cases than is the case when $R_s=0$, the latter being essentially that employed in prior determinations of Q's and admittances. In many cases, it has been the custom to assume that $R_s=0$. As will be seen, this can often lead to appreciable error.

CONSIDERATIONS REGARDING THE Q's

Since admittance measurements on the system of which Fig. 5 is the equivalent circuit will be made at points corresponding to 2–2′, R_s must be considered as one of the *internal* lossy elements. As a consequence, the Q's will be defined as follows:

$$Q_L=\frac{2\pi \text{ Energy stored in circuit}}{\text{Energy dissipated in } Y_0,\ R_s \text{ and } G \text{ per cycle}}, \quad (1)$$

$$Q_0=\frac{2\pi \text{ Energy stored in circuit}}{\text{Energy dissipated in } G \text{ and } R_s \text{ per cycle}}, \quad (2)$$

$$Q_{ext}=\frac{2\pi \text{ Energy stored in circuit}}{\text{Energy dissipated in } Y_0 \text{ per cycle}}. \quad (3)$$

Q_L can be alternatively expressed as the ratio $f_0/\Delta f$ where $f_0=1/2\pi(LC)^{\frac{1}{2}}$ and Δf is the difference in frequency between the two points on either side of resonance at which the total circuit susceptance as measured between any two points of the circuit is equal in magnitude to the conductance as measured between the same two points. The points 3–3′ of Fig. 5 are convenient points to use. If B is the susceptance measured across those points, then the frequencies used in the determination of Δf are those for which:

$$B=\pm\left[G+\frac{1}{R_s+1/Y_0}\right]. \quad (4)$$

If $R_s=0$, a simple graphical means exists for determining Q_L.[1] By means of a standing wave machine one measures the voltage standing wave ratio ($VSWR$)[3] as a function of frequency in the neighborhood of resonance. When plotted, this yields a curve of the form shown in Fig. 6 (marked $R_s=0$).

Furthermore (still restricting the discussion to the case where $R_s=0$), one can measure the admittance across terminals 2–2′ (or in this case, since $R_s=0$, across terminals 3–3′), by means of a standing wave machine. This admittance when plotted on a circular reflection chart yields a circular locus which passes through the infinity point. This is the circle marked $R_s=0$ in Fig. 7. One now finds the point of intersection between this circle and the 45° straight line A_∞ which passes through the infinity point. Designate the $VSWR$ corresponding to this point by σ_{L0}. Find the intersections f_1 and f_2 of the horizontal line through σ_{L0} on Fig. 6 with the $R_s=0$ curve. Then $\Delta f=f_2-f_1$ is the quantity which when divided into f_0 yields the desired Q_L. One proceeds in a similar fashion in the determination of Q_{ext} except that the $VSWR$ used is that corresponding to the intersection of the admittance locus with the line for which $B/Y_0=1$. This point is designated as σ_{00} in Fig. 7. When Q_L and Q_{ext} are known, Q_0 can be determined from the relation:

$$1/Q_L=1/Q_0+1/Q_{ext}. \quad (5)$$

The circuit efficiency can be obtained from the well-known relation:

$$\eta_e=1-(Q_L/Q_0)=Q_L/Q_{ext}. \quad (6)$$

Unfortunately, as can readily be seen, these simple geometrical methods fail when $R_s\neq 0$. In the latter case, the admittance locus across points 2–2′ of Fig. 5 when plotted on a reflection chart is still a circle, but one which does not pass through the infinity point.

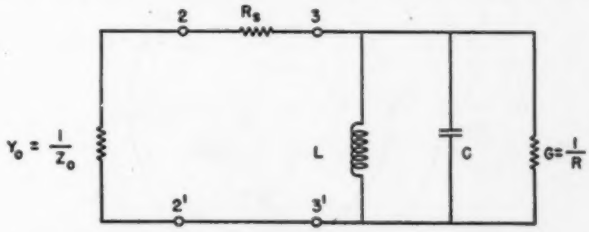


FIG. 5. Final equivalent circuit of parallel resonant circuit coupled into load through a lossy transformer.

[3] VSWR is the ratio of maximum to minimum "voltage" along the line.

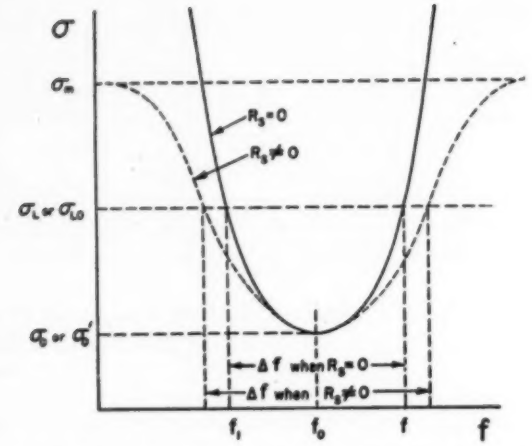


FIG. 6. Plot of *VSWR vs.* frequency for a resonant circuit coupled to an external load ($R_s=0$; $R_s\neq 0$).

Two such circles are shown in Fig. 7 and it is obvious at once that the method just described cannot work, since one of the loci does not even intersect the 45° line or the $B/Y_0=1$ curve. It would still be convenient, if possible, to set up a means whereby the various Q's and η_c can be obtained in a reasonably straightforward fashion from standing wave measurements. It is the purpose of this paper to show how this can be done.

DETERMINATION OF Q_L

There are many ways of proceeding to determine the desired Q's for the case where $R_s\neq 0$. The authors have felt that it would be desirable to set up a procedure which would parallel as closely as possible that for the case wherein $R_s=0$.

One proceeds as before to measure *VSWR* and admittance as a function of frequency, the measurements being made across points 2–2′ of Fig. 5. The *VSWR* curve will have the form shown in Fig. 6 (marked $R_s\neq 0$). It is seen that the *VSWR* does not increase indefinitely on either side of resonance but approaches a

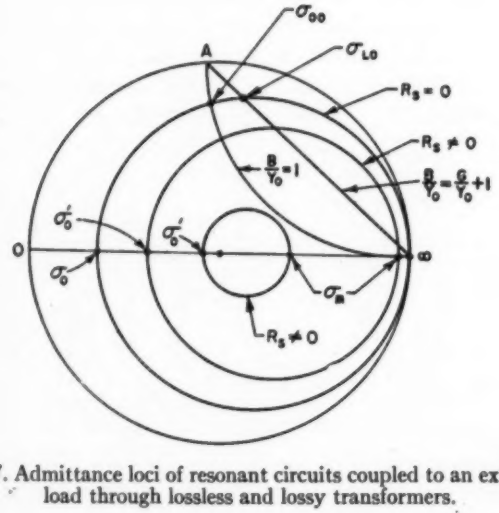


FIG. 7. Admittance loci of resonant circuits coupled to an external load through lossless and lossy transformers.

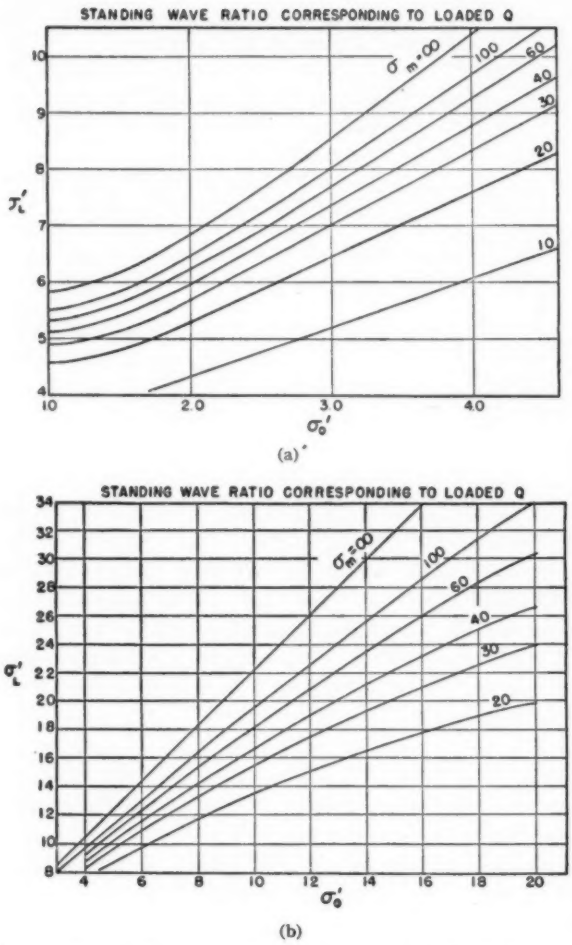


FIG. 8. Standing wave ratio corresponding to loaded Q—(a) Small σ_0'. (b) Large σ_0'.

limiting value σ_m which is given by:

$$\sigma_m=\frac{1}{R_sY_0}=\frac{Z_0}{R_s}. \tag{7}$$

The admittance locus at "zero" or "infinite" frequency does not pass through the infinity point but through a point on the real axis corresponding to $VSWR=\sigma_m$. Let the other intersection of the admittance locus with the real axis (the resonant point) correspond to $VSWR=\sigma_0'$. If the admittance locus encloses the origin, then:

$$\sigma_0'=Y_0\left[\frac{R_sG+1}{G}\right]=\frac{R_sG+1}{Z_0G}. \tag{8a}$$

This is denoted as the *overmatched* case. If the admittance locus does not enclose the origin, then:

$$\sigma_0'=\frac{1}{Y_0}\left[\frac{G}{R_sG+1}\right]=\frac{Z_0G}{R_sG+1}. \tag{8b}$$

This is denoted as the *undermatched* case. We must determine the value of *VSWR* present at 2–2′ of Fig. 5, corresponding to (4) which holds at 3–3′. This value of

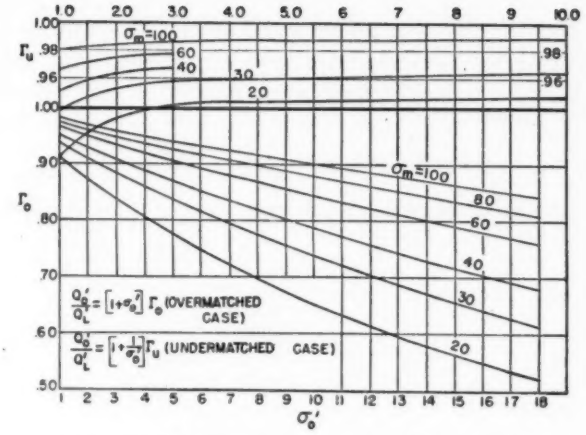


FIG. 9. Ratio factor Γ of unloaded to loaded Q in the presence of series losses.

VSWR which is needed for the determination of Q_L is denoted by σ_L.

Let:

$$Z_t=R_s+\frac{1}{G+jB} \tag{9}$$

be the impedance of the circuit to the right of 2–2′ in Fig. 5. Then the reflection coefficient at the plane through 2–2′ is given by:

$$\Gamma=-\frac{Z_0-Z_t}{Z_0+Z_t}=-\frac{Z_0G+jZ_0B-R_sG-jR_sB-1}{Z_0G+jZ_0B+R_sG+jR_sB+1} \tag{10}$$

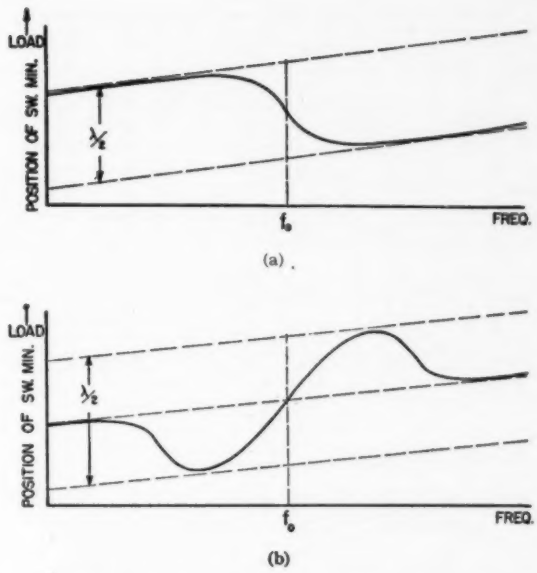


FIG. 10. Position of standing wave minimum as a function of frequency. (a) Resonant circuit overmatched to load. (b) Resonant circuit undermatched to load.

where:

$$Z_0=1/Y_0.$$

If (4) is substituted in (10) and the corresponding value of Γ is denoted by Γ_L the following expression is obtained:

$$\Gamma_L=-\frac{\left(1-\frac{R_sG+1}{Z_0G}\right)+j\left(1+\frac{2Z_0}{(R_s+Z_0)Z_0G}-\frac{R_sG+1}{Z_0G}\right)}{\left(1+\frac{R_sG+1}{Z_0G}\right)+j\left(1+\frac{R_sG+1}{Z_0G}\right)}. \tag{11}$$

The *VSWR* corresponding to Eq. (4) has been chosen to be denoted by σ_L. Then from the well-known relation between reflection coefficient and standing wave ratio,

$$\sigma_L=\frac{1+|\Gamma_L|}{1-|\Gamma_L|}. \tag{12}$$

The overmatched case will first be treated. Combining (7), (8a), (11), and (12), the following is obtained:

$$\sigma_L=\frac{(\sigma_m+1)(\sigma_0'+1)+[(\sigma_m-\sigma_0')^2+(\sigma_0'\sigma_m-1)^2]^{\frac{1}{2}}}{(\sigma_m+1)(\sigma_0'+1)-[(\sigma_m-\sigma_0')^2+(\sigma_0'\sigma_m-1)^2]^{\frac{1}{2}}}. \tag{13}$$

For the undermatched case (7), (8b), (11), and (12) are combined. The resultant expression is identical with that of (13). Thus (13) is the relation which enables σ_L to be determined from only two quantities, viz. σ_0' and σ_m. Thus for a determination of Q_L, it is unnecessary to know whether one is dealing with an undermatched or overmatched case. If $R_s=0$ then $\sigma_m=\infty$. In that case (13) reduces to:

$$\sigma_L=\frac{\sigma_0+1+(\sigma_0^2+1)^{\frac{1}{2}}}{\sigma_0+1-(\sigma_0^2+1)^{\frac{1}{2}}} \tag{14}$$

which checks with the expression of Lawson.[1]

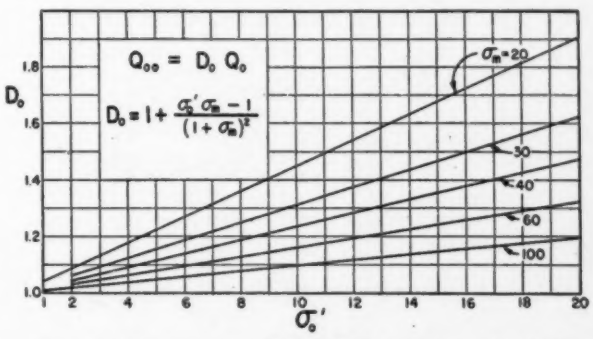


FIG. 11. Ratio factor D_0 of unloaded Q in absence of series losses, to unloaded Q in presence of series losses, for the overmatched case.

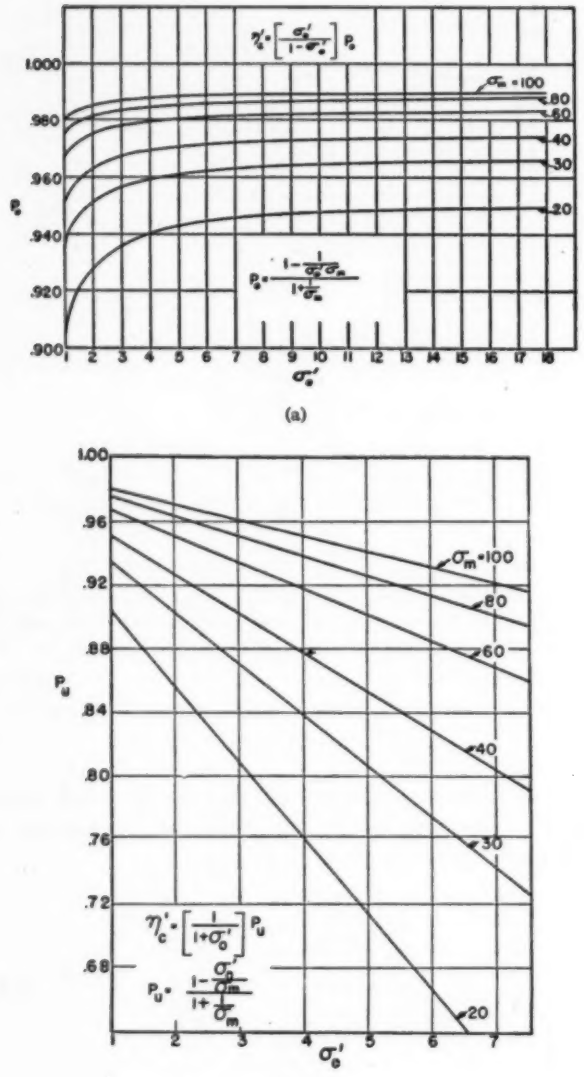


FIG. 12. Circuit efficiency correction factor. (a) Overmatched case. (b) Undermatched case.

Find the intersection of the horizontal line through σ_L in Fig. 6 with the curve marked $R_s \neq 0$. This yields the values f_1' and f_2' from which one obtains the value of Q_L from the relation $Q_L = f_0/\Delta f$, where $\Delta f = f_2' - f_1'$.

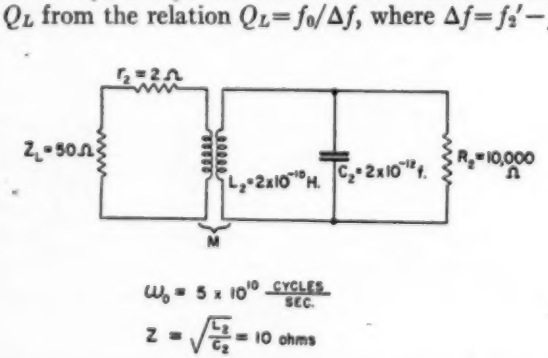


FIG. 13. A parallel resonant circuit coupled to an external load through a lossy transformer.

σ_L as a function of σ_0' for various values of σ_m is plotted in Figs. 8a and 8b. The curve marked $\sigma_m = \infty$ is the one corresponding to $R_s = 0$, i.e., the case of no series losses.

DETERMINATION OF Q_0 AND Q_{ext}

Let V be the peak r-f voltage across C in Fig. 5. Then (2) can be written as:

$$Q_0 = \frac{\frac{1}{2}\omega_0 C V^2}{\frac{1}{2}GV^2 + \frac{1}{2}\frac{V^2}{R_s}\left(\frac{R_s}{R_s+Z_0}\right)^2} = \frac{\omega_0 CR}{1+\dfrac{1}{GR_s+GZ_0\left(\dfrac{Z_0}{R_s}+2\right)}}. \tag{15}$$

In view of (7), (15) can be rewritten as:

$$Q_0 = \frac{\omega_0 CR}{1+[\sigma_m/GZ_0(1+\sigma_m)^2]}. \tag{16}$$

By proceeding in the same fashion (1) can be put in the form:

$$Q_L = \frac{\omega_0 CR}{1+\dfrac{\sigma_m}{GZ_0(1+\sigma_m)}}. \tag{17}$$

From (16) and (17) the following is obtained:

$$Q_0 = \left[\frac{GZ_0(1+\sigma_m)^2+\sigma_m(1+\sigma_m)}{GZ_0(1+\sigma_m)^2+\sigma_m}\right]Q_L. \tag{18}$$

For the overmatched case from (8a):

$$\frac{1}{GZ_0} = \frac{Y_0}{G} = \sigma_0' - Y_0R_s = \sigma_0' - \frac{1}{\sigma_m}. \tag{19}$$

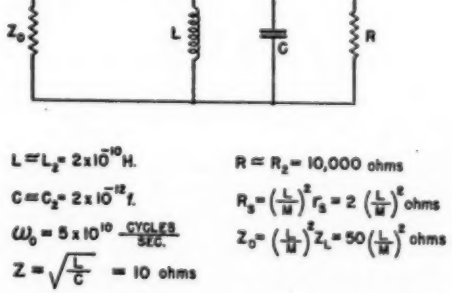


FIG. 14. Equivalent circuit of a parallel resonant circuit coupled to an external load through a lossy transformer.

TABLE I. Values of P as a function of σ_0' and σ_m.

	σ_0'	σ_m 100	80	60	40	30	20
Overmatched case	1	.9801	.9753	.9672	.9512	.9355	.9048
	2	.9851	.9814	.9754	.9634	.9516	.9285
	3	.9867	.9835	.9781	.9675	.9569	.9365
	4	.9876	.9845	.9795	.9695	.9596	.9404
	5	.9881	.9851	.9803	.9707	.9612	.9428
	6	.9884	.9855	.9808	.9715	.9623	.9444
	8	.9888	.9861	.9815	.9725	.9637	.9464
	10	.9891	.9864	.9819	.9731	.9645	.9476
	12	.9892	.9866	.9822	.9735	.9650	.9484
	14	.9893	.9867	.9824	.9738	.9654	.9489
	16	.9894	.9868	.9825	.9740	.9657	.9494
	18	.9895	.9869	.9826	.9742	.9659	.9497
	20	.9896	.9870	.9827	.9743	.9661	.9500
Undermatched case	2	.9702	.9629	.9508	.9262	.9032	.8571
	3	.9603	.9506	.9344	.9024	.8709	.8095
	4	.9504	.9382	.9180	.8780	.8387	.7619
	5	.9405	.9259	.9016	.8536	.8064	.7142

Substitution of (19) in (18) yields:

$$Q_0=(1+\sigma_0')\Gamma_0 Q_L \tag{20}$$

where:

$$\Gamma_0=\frac{\sigma_m+1}{\sigma_m+\sigma_0'+2}. \tag{21}$$

For the *undermatched case*, the corresponding expressions are:

$$Q_0=\left(1+\frac{1}{\sigma_0'}\right)\Gamma_u Q_L \tag{22}$$

where:

$$\Gamma_u=\frac{\sigma_m+1}{\sigma_m+\dfrac{1}{\sigma_0'}+2}. \tag{23}$$

When $R_s=0$, $\sigma_m=\infty$, and $\Gamma_0=\Gamma_u=1$, (20) and (22) then reduce to the well-known expressions for the case of no series losses.

Γ_0 and Γ_u as functions of σ_0' for various values of σ_m are plotted in Fig. 9. Thus, from: 1—the value of Q_L determined in the manner shown in the previous section, 2—the values of σ_0' and σ_m, and 3—a knowledge as to whether one is dealing with the undermatched or overmatched case, one can determine Q_0. It is thus necessary to determine which of the two possible cases one is dealing with. While this can be done by plotting the admittance locus and seeing whether or not it encloses the origin, this is actually unnecessary. If the position of the standing wave minimum is plotted as a function of frequency, curves of the form of Figs. 10a or 10b are obtained. The former distinguishes the overmatched case and the latter the undermatched case. Thus to distinguish between the two cases it is merely necessary to observe whether or not the standing wave minimum experiences a reversal as one approaches and goes through resonance.

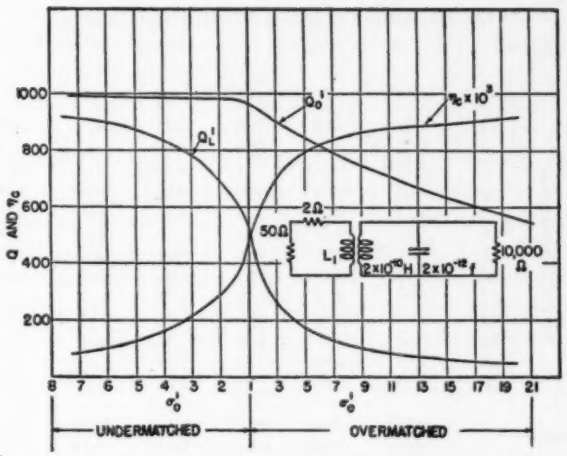


FIG. 15. Plots of loaded and unloaded Q's and of circuit efficiency for the circuit of Fig. 13.

Finally, Q_{ext} may be determined from Q_L and Q_0 by means of (5).

DETERMINATION OF Q_{00}

Q_{00} is defined as being given by $\omega_0 CR$ where C and R are elements of Fig. 5. Q_{00} is thus a measure of the losses in the resonant circuit proper, distinct from those due to the coupling element. Thus, a knowledge of both Q_0 and Q_{00} enables one to assign the losses to their proper place. In practice this may be useful in indicating a point of faulty construction or design. From (16) and (7), (8a), and (19) one obtains for the overmatched case:

$$Q_{00}=D_0 Q_0 \tag{24}$$

where:

$$D_0=1+\frac{\sigma_0'\sigma_m-1}{(1+\sigma_m)^2}. \tag{25}$$

For the undermatched case, the corresponding expres-

TABLE II. Values of Γ as a function of σ_0' and σ_m.

	σ_0'	σ_m 100	80	60	40	30	20
Overmatched case	1.0	.981	.976	.968	.953	.939	.913
	2.0	.971	.964	.953	.932	.912	.875
	3.0	.962	.953	.938	.911	.886	.840
	4.0	.953	.942	.924	.891	.861	.808
	5.0	.944	.931	.910	.872	.838	.778
	6.0	.935	.920	.897	.854	.816	.750
	8.0	.918	.900	.871	.820	.775	.700
	10.0	.902	.880	.847	.788	.738	.656
	12.0	.886	.862	.824	.759	.705	.618
	14.0	.871	.844	.803	.732	.674	.583
	16.0	.856	.827	.782	.707	.646	.553
	18.0	.842	.810	.763	.683	.620	.525
	20.0	.828	.794	.744	.661	.596	.500
Undermatched case	2.0	.985	.982	.976	.965	.954	.933
	3.0	.987	.984	.979	.969	.959	.940
	4.0	.988		.980		.961	.944
	10.0	.989				.966	.950

TABLE III. Values of Φ as a function of σ_0' and σ_m.

	σ_0'	σ_m 100	80	60	40	30	20
Overmatched case	1	.990	.988	.984	.977	.971	.959
	2	.981	.976	.969	.955	.942	.919
	3	.971	.965	.954	.934	.915	.882
	4	.962	.953	.940	.913	.890	.848
	5	.953	.943	.926	.894	.866	.817
	6	.945	.932	.912	.876	.843	.788
	8	.927	.911	.886	.841	.801	.735
	10	.911	.891	.861	.808	.763	.689
	12	.895	.872	.838	.778	.728	.649
	14	.879	.854	.816	.751	.696	.613
	16	.864	.837	.795	.725	.667	.580
	18	.850	.820	.775	.701	.641	.551
	20	.836	.804	.756	.678	.616	.525
Undermatched case	2	.9952	.9940	.9923	.9889	.9855	.9800
	3	.9968	.9960	.9949	.9928	.9906	.9873
	4	.9976	.9970	.9963	.9947	.9931	.9910
	5	.9981	.9976	.9971	.9959	.9947	.9932

TABLE IV. Values of σ_L' as function of σ_0' with σ_m as parameter.

σ_0'	σ_m ∞	100	80	60	40	30	20	10
1.0	5.83	5.50	5.44	5.32	5.12	4.90	4.56	3.74
1.4	6.06	5.73	5.66	5.54	5.31	5.05	4.73	3.89
1.7	6.42	6.06	5.99	5.85	5.61	5.32	4.97	4.09
2.0	6.85	6.46	6.38	6.23	5.97	5.71	5.30	4.33
2.4	7.50	7.09	6.95	6.80	6.49		5.73	4.67
2.7	8.01	7.52	7.42	7.24	6.89		6.10	4.94
3.0	8.55	8.01	7.89	7.70	7.32	7.00	6.43	5.20
3.4	9.28	8.67	8.53	8.30	7.91		6.90	5.58
3.7	9.84	9.17	9.02	8.78	8.33		7.26	5.81
4.0	10.40	9.67	9.51	9.25	8.76	8.34	7.61	6.06
4.5	11.36	10.51	10.34	10.04	9.49		8.17	6.48
5.0	12.32	11.36	11.15	10.81	10.21	9.66	8.73	6.88
6.0	14.27	13.04	12.78	12.34	11.59	10.92	9.81	7.62
7.0	16.22	14.71	14.37	13.86	12.96	12.09	10.83	8.29
8.0	18.19	16.35	15.95	15.33	14.25	13.27	11.81	8.91
9.0	20.16	17.97	17.51	16.78	15.50	14.42	12.68	9.48
10.0	22.16	19.57	19.05	18.17	16.72	15.51	13.57	
12.0	26.11	22.70	21.99	20.91	19.03	17.50	15.15	
14.0	30.09	25.70	24.82	23.50	21.20	19.36	16.50	
16.0	34.09	28.63	27.57	25.94	23.24	21.08	17.79	
18.0	38.05	31.47	30.19	28.41	25.10	22.67	18.95	
20.0	41.90	34.20	32.72	30.44	26.86	24.15		

sions are:

$$Q_{00}=D_uQ_0, \tag{26}$$

where:

$$D_u=1+\frac{\sigma_m-\sigma_0'}{\sigma_0'(1+\sigma_m)^2}. \tag{27}$$

D_0 is plotted in Fig. 11 as a function of σ_0' for various values of σ_m. D_u differs so slightly from unity for all usual combinations of σ_0' and σ_m that its values are not plotted.

CIRCUIT EFFICIENCY

From (6) the circuit efficiency yields:

$$\eta_c=1-(Q_L/Q_0). \tag{28}$$

Substitution of (20) and (22) into (28) yields:

$$\eta_c=(\sigma_0'/\sigma_0'+1)P_0 \tag{29}$$

for the overmatched case, and

$$\eta_c=(1/\sigma_0'+1)P_u \tag{30}$$

for the undermatched case, where:

$$P_0=\frac{1-(1/\sigma_0'\sigma_m)}{1+(1/\sigma_m)}, \tag{31}$$

$$P_u=\frac{1-(\sigma_0'/\sigma_m)}{1+(1/\sigma_m)}. \tag{32}$$

When $R_s=0$, $P_0=P_u=1$, and (29) and (30) reduce to the well-known expressions for the case wherein there are no series losses. P_0 and P_u as functions of σ_0' for various values of σ_m are plotted in Figs. 12a and 12b.

SUMMARY OF EXPERIMENTAL PROCEDURES

The determination of the Q's and η_c can be determined in a sequence of steps which proceed as follows:

1. By means of a standing wave machine one measures *VSWR* as a function of frequency and plots a curve of the form shown in Fig. 6. From this curve one determines σ_0' and σ_m.
2. By observing the manner in which the standing wave minimum behaves as one goes through resonance, one determines whether one is dealing with an undermatched or overmatched case. (See Figs. 10a and 10b.)
3. Knowing σ_0' and σ_m one determines σ_L from Fig. 8.
4. From σ_L one determines Q_L from Fig. 6 by means of the relation $Q_L=f_0/\Delta f$.
5. One now obtains Q_0 by means of (20) or (22), the knowledge of Q_L and the value of Γ_0 or Γ_u obtained from Fig. 9.
6. Q_{ext} is now determined from (5).
7. Q_{00} is determined from (24) or (26) and from the knowledge of Q_0 and from D_0 or D_u obtained from Fig. 11.
8. Finally, η_c is determined from (29) or (30) using the values of P_0 or P_u from Figs. 12a and 12b.

ILLUSTRATIVE CASE

In order to depict the manner in which the neglect of series losses can introduce appreciable error, the behavior of a typical case as illustrated in Fig. 13 or its equivalent as given in Fig. 14 will be studied.

Let:

$$(L/M)^2=2000$$
$$R_s=4000 \text{ ohms.}$$
$$Z_0=100{,}000 \text{ ohms.}$$

Then:

$$\sigma_m=25$$

$$\sigma_0'=\frac{Z_0}{R_s+R}=\frac{100{,}000}{14{,}000}=7.14.$$

Q_L is obtained by means of (17) and the value $Q_L=912$.

Then from (22) (since this is obviously an undermatched case) and Fig. 9 the value $Q_0=992$ is obtained.

From (28) $\eta_c = 0.0844$ is obtained. Proceeding in the same fashion, the Q's and the circuit efficiency were computed over a wide coupling range.

The results are plotted in Fig. 15 as a function of σ_0'. An examination of this figure shows at once that the common assumption that Q_0 may be considered a constant, is seriously in error for cases where the series resistance is sufficiently great to make the *VSWR* far from resonance differ appreciably from infinity.

While Q_0 is not a constant, Q_{00} (the Q associated with the cavity under measurement, uninfluenced by the series losses) is a constant, as it should be, and is given by $\omega_0 CR = 1000$.

SUMMARY

The presence of series losses in microwave circuits makes itself felt in impedance measurements carried through by standing wave machine techniques, in that the *VSWR* far from resonance is no longer infinite. Neglect of the effect of series losses results in errors in determinations of the Q's and the circuit efficiency. The Q determinations have their greatest error for grossly overmatched systems, whereas the circuit efficiencies are in greatest error for lightly loaded systems (severely undermatched cases). Unless series losses are quite negligible the internal or unloaded $Q(Q_0)$ cannot be considered as a constant independent of coupling.

A family of curves is included which enables the various Q's and the circuit efficiency to be determined from the following three items:

1. *VSWR* at resonance (σ_0');
2. *VSWR* far from resonance (σ_m);
3. Knowledge as to whether the case is "*undermatched*" or "*overmatched*."

The Application of Non-Integral Legendre Functions to Potential Problems

R. N. Hall

Research Laboratory, General Electric Company, Schenectady, New York

(Received February 14, 1949)

There are a number of important potential problems having axial symmetry and involving a conical boundary whose solutions require the use of zero-order Legendre functions of real but non-integral degree. The numerical evaluation of such problems is relatively simple once the zeros of these functions have been determined and when certain integrals have been evaluated. This paper presents tables and curves giving this information, together with a number of approximate expressions for these quantities. A number of potential distributions involving conducting cones, spheres, and rings are worked out to illustrate the application of these functions.

I. INTRODUCTION

It is well known that solutions of Laplace's equation, when expressed in spherical polar coordinates, usually involve Legendre functions of the polar angle. When the polar axis is not excluded from the region of space under consideration, these functions generally take the form of Legendre polynomials whose properties have been thoroughly investigated. There is, however, a large class of problems involving a conducting cone as one boundary whose solutions require the use of Legendre functions of real, but non-integral degree. A discussion of such problems may be found in Smythe,[1] and Hobson,[2] page 444. The solutions are generally straightforward in principle, but the numerical results are not easily obtained because the behavior of these functions and their integral properties do not appear to have been worked out in sufficient detail, and because of the lack of tabulated values of the functions.

The purpose of this paper is to extend the knowledge of the behavior of the Legendre functions of zero order and real non-integral degree, and to include tables and curves which permit the evaluation of a number of useful potential distributions. Special emphasis is placed upon the behavior of the functions of nearly integral degree which are involved in potential distributions about needle points and spheres supported by thin wires.

These functions may be used to solve a number of problems having axial symmetry such as those illustrated in Fig. 1. The ring and cone geometry shown in Fig. 1d leads to a close approximation to the conical point-to-plane potential distribution, since by a proper choice of angle α, an equipotential surface may be found which is nearly plane over a large area near the tip of the cone. The effect of a lead-in wire to a spherical condenser may be investigated by the electrode configuration illustrated in Fig. 1e. This geometry is frequently encountered in photoelectric studies[3] and in secondary electron emission work.[4] These problems are considered in detail in Part VI as illustrative examples. The study of radiation from conical antennas as treated by Schelkunoff[5] is another example of a problem whose solution involves non-integral Legendre functions.

[1] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939), Section 5.25.

[2] E. W. Hobson, *Spherical and Ellipsoidal Harmonics* (Cambridge University Press, London, 1931).

[3] Apker, Taft, and Dickey, Phys. Rev. **74**, 1462 (1948).

[4] H. Bruining, *Die Sekundär-Elektronen-Emission fester Korper* (Verlag. Julius Springer, Berlin, 1942), p. 6.

[5] S. A. Schelkunoff, Proc. I.R.E. **29**, 493 (1941).

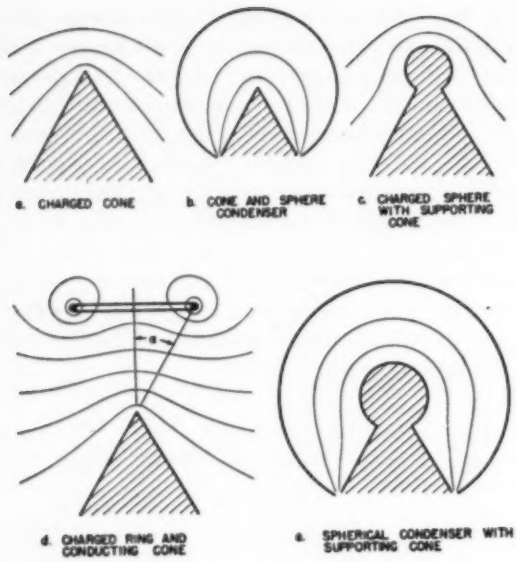

a. CHARGED CONE
b. CONE AND SPHERE CONDENSER
c. CHARGED SPHERE WITH SUPPORTING CONE
d. CHARGED RING AND CONDUCTING CONE
e. SPHERICAL CONDENSER WITH SUPPORTING CONE

FIG. 1. Axially symmetric potential distributions which may be expressed in terms of non-integral Legendre functions.

II. METHOD OF SOLUTION

For potential distributions having axial symmetry, a general solution may be expressed in spherical polar coordinates in the form:

$$V=\sum[A_\nu P_\nu(\cos\theta)+B_\nu Q_\nu(\cos\theta)](r^\nu+C_\nu r^{-\nu-1}).$$

P_ν and Q_ν are Legendre functions of the first and second kinds and the index ν is summed over the proper set of values. When the potential is known to be finite along the axis $\theta=0$ and π, the coefficients B_ν vanish and the index ν takes on positive integral values. If the half of the axis at $\theta=\pi$ is excluded by a conducting cone at $\theta=\theta_0$, a different method of solution is required. One method is to take the cone to be at zero potential and to sum over all indices ν which cause the Legendre function to vanish on the cone. The remaining coefficients are then chosen to satisfy the rest of the boundary conditions.

In this method, the summation index ν takes on successive values ν_s which in general are real but non-integral. The functions $P_\nu(\cos\theta)$ have a logarithmic singularity at $\theta=\pi$, so the other linearly independent solutions $P_\nu(-\cos\theta)$ are excluded by the requirement of finiteness along the axis $\theta=0$. The potential distribution may, therefore, be written in the form

$$V=\sum_s P_{\nu_s}(\cos\theta)[A_s r^{\nu_s}+B_s r^{-\nu_s-1}] \qquad (1)$$

where $P_{\nu_s}(\cos\theta_0)=0$.

The evaluation of the constants A_s and B_s generally involves integrals of the form

$$I_s(\cos\theta_0)=\int_0^{\theta_0} P_{\nu_s}(\cos\theta)\sin\theta d\theta$$

$$H_s(\cos\theta_0)=\int_0^{\theta_0}[P_{\nu_s}(\cos\theta)]^2\sin\theta d\theta.$$

The solution of cone problems with axial symmetry requires, therefore, a knowledge of the location of the zeros of the non-integral Legendre functions together with the evaluation of the integrals I_s and H_s.

III. LEGENDRE FUNCTIONS OF NON-INTEGRAL DEGREE

Hobson has shown[2] that Legendre functions of zero order and unrestricted degree are given by the hypergeometric function

$$P_\nu(\mu)=F(-\nu,\nu+1;1;(1-\mu)/2) \qquad (2)$$

where $\mu=\cos\theta$. This function has a logarithmic singularity at $\mu=-1$ except in the special cases when ν is an integer. When ν is not an integer, a complete solution to Legendre's equation is evidently given by a linear combination of $P_\nu(\mu)$ and $P_\nu(-\mu)$. The latter is excluded when the potential is required to be finite at $\theta=0$. Furthermore, $P_\nu(\mu)=P_{-\nu-1}(\mu)$ so it is not necessary to consider values of ν less than $-\frac{1}{2}$. We therefore seek successive values of ν in the region $\nu\geqslant-\frac{1}{2}$ which make $P_\nu(\mu_0)$ vanish on the surface of the cone $\mu_0\equiv\cos\theta_0$.

The behavior of Legendre functions of non-integral degree is illustrated in Fig. 2. When ν lies between the integers $n-1$ and n, $P_\nu(\mu)$ has n zeros in the range $-1\leqslant\mu\leqslant1$ and approaches plus or minus infinity as $\mu\to-1$ according to whether n is an even or an odd integer. Since Eq. (2) converges very slowly near $\mu=-1$, practical considerations limit its usefulness to the region of positive μ. An expression which is useful when μ is negative is given by Hobson[2]

$$P_\nu(\mu)=\frac{\sin\nu\pi}{\pi}\left\{[\log(1+\mu)/2+2C+\psi(\nu+1)+\psi(-\nu)]P_\nu(-\mu)+\sum_{r=1}^{\infty}\frac{(-1)^r\Gamma(\nu+r+1)}{\Gamma(\nu-r+1)(r!)^2}\varphi(\nu,r)\left(\frac{1+\mu}{2}\right)^r\right\} \qquad (3)$$

where

$$\psi(\nu)=\frac{d}{d\nu}\log\Gamma(\nu),$$

$$C=\text{Euler's constant}=0.5772\cdots,$$

$$\varphi(\nu,r)=\left[\frac{1}{\nu+1}+\frac{1}{\nu+2}+\cdots+\frac{1}{\nu+r}\right]+\left[\frac{1}{-\nu}+\frac{1}{-\nu+1}+\cdots+\frac{1}{-\nu+r-1}\right]-2\left[\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{r}\right].$$

Davis[6] gives tables of the gamma-function and its

[6] H. T. Davis, *Tables of the Higher Mathematical Functions* (Principia Press, Inc., Bloomington, Indiana, 1933, 1935), Vols. 1 and 2.

logarithmic derivatives which are useful in the evaluation of Eq. (3) and other expressions derived from it.

Snow[7] and Schelkunoff[5] give two other somewhat similar expressions which are less convenient, but which can be shown to reduce to Hobson's equation. Equation (3) converges rapidly near $\mu=-1$ and is useful in the region $-1<\mu<0$ when ν is not large compared with unity.

Schelkunoff gives approximate expressions for Legendre functions of nearly integral degree:

$$P_\epsilon \approx 1+\epsilon \log(\mu+1)/2$$
$$P_{1+\epsilon} \approx \mu+\epsilon[\mu-1+\mu \log(\mu+1)/2].$$

These expressions represent the Legendre functions over the entire range of μ with an error which is of the order of ϵ^2. The recurrence formulas may be used to generate the higher nearly integral functions.

It is often useful to note that the half-integral Legendre functions may be expressed in terms of complete elliptic integrals of modulus k where $k^2=(1-\mu)/2$.

$$P_{-\frac{1}{2}}(\mu)=(2/\pi)K$$
$$P_{\frac{1}{2}}(\mu)=(2/\pi)(2E-K)$$

and so on.

IV. ZEROS

The zero-order Legendre functions defined by Eq. (2) may be regarded as functions of the two variables, ν and μ. Hobson has shown that for any given real values of μ in the range $-1\leqslant\mu\leqslant 1$, the Legendre functions vanish only for certain real values of ν. As previously mentioned, we need only consider values of ν greater than $-\frac{1}{2}$. With the argument held constant at $\mu=\mu_0$, the Legendre functions vanish at successive values of the degree ν_s, $s=0, 1, 2\cdots$, which we will order in such a way that $\nu_0<\nu_1<\nu_2\cdots$. On the other hand, we may regard the argument μ as the independent variable with the degree ν held constant. With s as defined above, the function $P_{\nu_s}(\mu)$ vanishes s times between $\mu=1$ and the zero at $\mu=\mu_0$. The zeros may therefore be mapped on the $\mu-\nu$-plane by a family of curves corresponding to successive values of s.

In a given problem, the cone angle θ_0 remains fixed and we wish to determine successive roots ν_s of the equation $P_\nu(\cos\theta_0)=0$, where $P_\nu(\mu_0)$ is regarded as a function of the index ν. The inset in Fig. 3 illustrates the variation of the roots ν_s with the cone angle. Except when μ_0 is close to -1, as is the case with sharp needle-like cones, ν_s varies slowly with μ_0 and the successive values of ν_s may be determined from the zeros of the integral and half-integral functions by the use of interpolation formulas. These zeros are given in Table I.

Since problems involving thin conducting cones ($\mu_0\approx -1$) are of considerable importance, we shall pay special attention to this region. Inspection of Eq. (3) shows that in this case, the sum may be neglected and the zeros are determined by the vanishing of the expression in the bracket. By making use of the properties of the psi-function, this expression may be reduced to

$$\log(\mu+1)/2+2C+2\psi(\nu+1)+\pi \cot\pi\nu=0. \quad (4)$$

Greater accuracy may be obtained by keeping first-order terms in $(1+\mu)$, thereby adding the quantity $(\frac{1}{2}+\nu+\nu^2)(1+\mu)$ to the left side of (4).

When $1+\mu$ is extremely small, ν_s takes on values

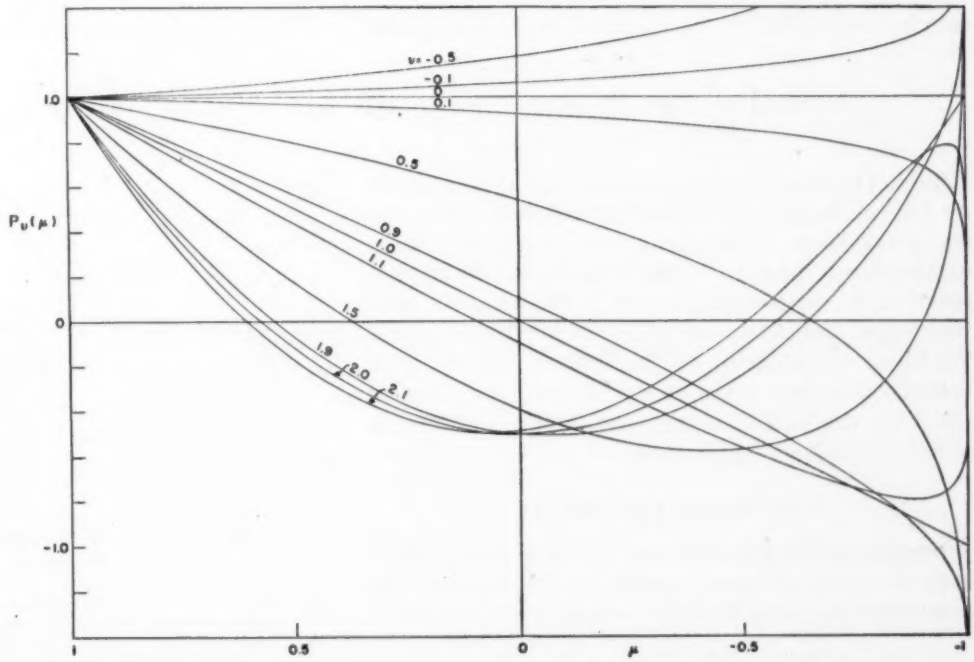


FIG. 2. Legendre functions of integral and non-integral degree.

[7] Chester Snow, "The hypergeometric and Legendre functions with applications to integral equations of potential theory" (National Bureau of Standards, 1942).

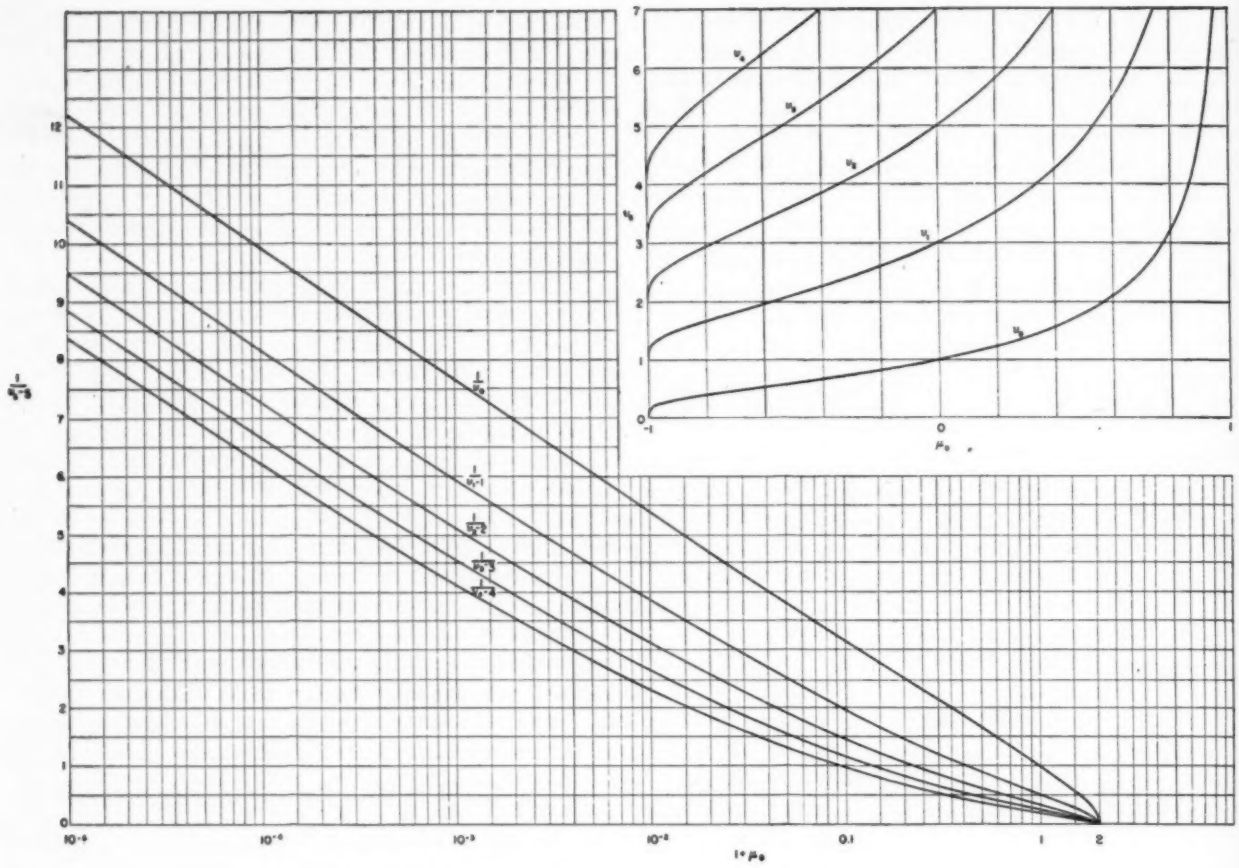


FIG. 3. Zeros of Legendre functions of unrestricted degree.

slightly greater than integers: $\nu_s = s + \epsilon_s$, $s = 0, 1, 2 \cdots$ where ϵ_s is a small positive quantity. In this case, it is readily shown that Eq. (4) leads to the limiting values

$$\frac{1}{\epsilon_s} = \frac{1}{\nu_s - s} = \log\left(\frac{2}{1+\mu_0}\right) - 2\left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{s}\right). \quad (5)$$

Table II gives a number of values of ν_s as determined by Eq. (4), including first- and second-order corrections where necessary. These results are plotted in Fig. 3, making use of the roots of the integral and half-integral functions to complete the curves in the region of large cone angle. The reciprocal of $(\nu_s - s)$ is plotted against $\log(1+\mu_0)$ as suggested by the form of Eq. (5). It is believed that the curves of Fig. 3 locate the zeros of the Legendre functions with an accuracy which is sufficient for most applications.

V. INTEGRAL PROPERTIES

Integrals of the Legendre functions and their squares over the region of space outside the conducting cone arise when use is made of the orthogonality relation

$$\int_{\mu_0}^{1} P_{\nu_s}(\mu) P_{\nu_t}(\mu) d\mu = H_s(\mu_0) \delta_{st} \quad (6)$$

to evaluate the coefficients of a potential distribution such as that of Eq. (1).

Making use of the recurrence formulas, we readily obtain

$$I_s(\mu_0) \equiv \int_{\mu_0}^{1} P_\nu(\mu) d\mu$$

$$= \frac{1}{2\nu+1}[P_{\nu-1}(\mu_0) - P_{\nu+1}(\mu_0)] \quad (7)$$

$$= \frac{P_{\nu-1}(\mu_0)}{\nu+1}. \quad (8)$$

In the above equation and in those that follow in this

TABLE I. Zeros of the Legendre functions of integral and half-integral degree.

ν	μ_0
1/2	−0.6522
1	0.00000
3/2	0.3724, −0.9038
2	±0.57735
5/2	0.6983, −0.2555, −0.9564
3	±0.77460, 0.00000
4	±0.86114, ±0.33998
5	±0.90618, ±0.53847, 0.00000
6	±0.93247, ±0.66121, ±0.23862

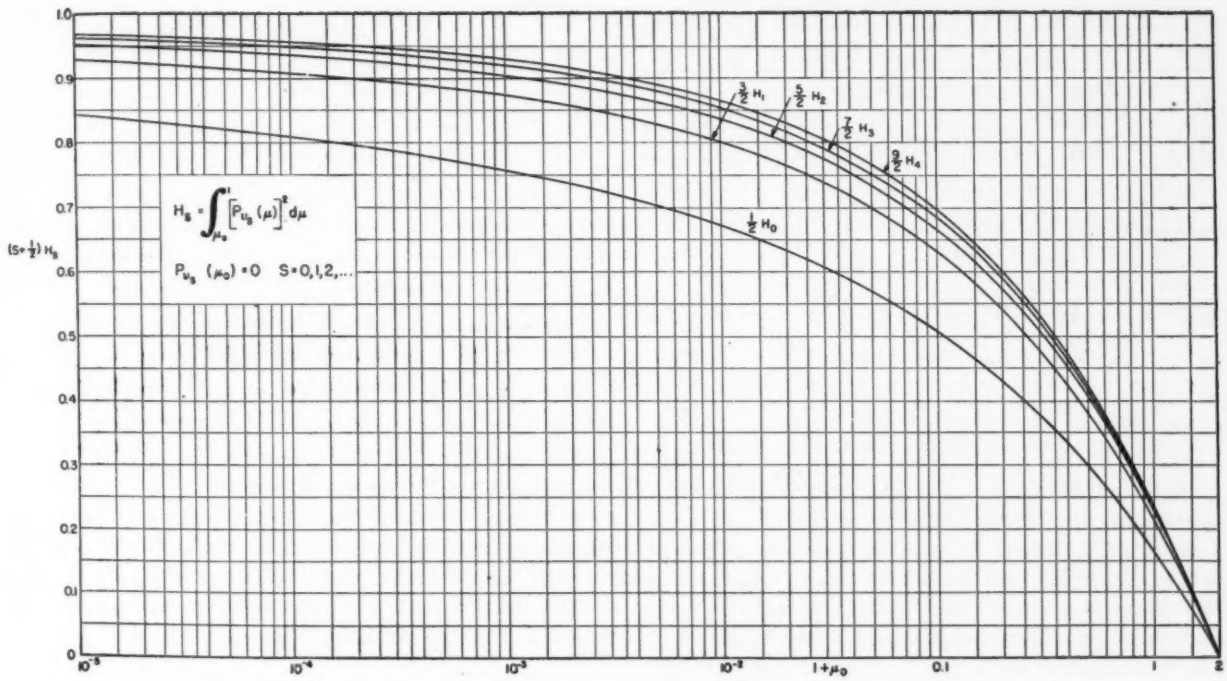


FIG. 4. Integrals of the squares of Legendre functions from $\mu=1$ to the $(s+1)$th zero located at $\mu=\mu_0$.

section, the subscript s is dropped from the index ν, and it is understood that ν is one of the set of values which makes $P(\mu)$ vanish when $\mu=\mu_0$.

When $1+\mu$ is not small, Eq. (8) proves convenient for the evaluation of I_s. An approximate expression which applies in the region where $1+\mu$ is small is obtained by substitution of (3) into (7), keeping first order terms in $(1+\mu)$. Making use of the psi-function relations, we obtain

$$I_\nu \approx \frac{\sin\nu\pi}{\pi}\left\{\frac{2}{\nu(\nu+1)}+(1+\mu_0)\times\left[1-\left(\log\frac{1+\mu}{2}+2C+\psi(\nu+1)+\psi(-\nu)\right)\right]\right\}.$$

The quantity in the bracket differs from unity by an amount which is of the order of $(1+\mu)$ as can be seen by the fact that μ_0 is a root of Eq. (3). We therefore obtain

$$I_\nu \approx \frac{\sin\nu\pi}{\pi}\left\{\frac{2}{\nu(\nu+1)}+(1+\mu_0)\right\}. \quad (9)$$

TABLE II. Zeros of Legendre functions near $\mu=-1$.

$1+\mu_0$	ν_0	ν_1	ν_2	ν_3	ν_4
0.01	0.18598	1.2598	2.3212	—	—
0.001	0.13086	1.1679	2.1959	3.2199	4.2421
0.0001	0.10072	1.1226	2.1378	3.1503	4.1612
0.00001	0.08182	1.0961	2.1053	3.1131	4.1193

Smythe[1] gives an expression from which we may calculate the integral of the square of the Legendre functions where $1+\mu$ is small

$$H_s(\mu_0) \equiv \int_{\mu_0}^{1} [P_\nu(\mu)]^2 d\mu = -\frac{(1-\mu_0^2)}{2\nu+1}\left(\frac{\partial P_\nu(\mu)}{\partial\mu}\frac{\partial P_\nu(\mu)}{\partial\nu}\right)_{\mu=\mu_0}. \quad (10)$$

Using the recurrence relations, there results

$$\left.\frac{\partial P_\nu(\mu)}{\partial\mu}\right|_{\mu=\mu_0} = \frac{\nu}{1-\mu_0^2}P_{\nu-1}(\mu_0) = \frac{\nu(\nu+1)}{1-\mu_0^2}I_s(\mu_0).$$

Keeping zero- and first-order terms in $(1+\mu)$ in Eq. (3), we obtain

$$\left.\frac{\partial P_\nu(\mu)}{\partial\nu}\right|_{\mu=\mu_0} \approx -\frac{\sin\nu\pi}{\pi}\left\{\left[\frac{\pi^2}{\sin^2\nu\pi}-2\psi'(\nu+1)\right]\times\left[1-\frac{\nu(\nu+1)}{2}(1+\mu_0)\right]-(2\nu+1)(1+\mu_0)\right\}.$$

The notation $\psi'(\nu+1)$ indicates the first derivative of the psi-function with respect to its argument.

Introducing these last two equations and (9) into (10) gives an expression for $H_s(\mu_0)$ which neglects

TABLE III. Integrals of squares of Legendre functions.

$1+\mu_0$	H_0	H_1	H_2	H_3	H_4
0.01	1.3414	0.5333	0.3342	—	—
0.001	1.5144	0.5816	0.3617	0.2628	0.2065
0.0001	1.6178	0.6058	0.3740	0.2708	0.2124
0.00001	1.6856	0.6197	0.3807	0.2749	0.2152

TABLE IV. Integrals of Legendre functions divided by integrals of their squares.

$1+\mu_0$	I_0/H_0	I_1/H_1	I_2/H_2	I_3/H_3	I_4/H_4
0.01	1.1882	−0.3100	0.2170	—	—
0.001	1.1353	−0.2178	0.1453	−0.1144	0.0966
0.0001	1.1045	−0.1657	0.1065	−0.0818	0.0677
0.00001	1.0848	−0.1330	0.0831	−0.0629	0.0513

quantities of the order of $(1+\mu_0)^2$:

$$H_s(\mu_0)\approx\frac{2}{2\nu+1}\left\{1-\frac{2\sin^2\nu\pi}{\pi^2}\psi'(\nu+1)-(2\nu+1)\frac{\sin^2\nu\pi}{\pi^2}(1+\mu_0)\right\}.$$

It may be noted that as $\mu\to-1$, ν takes on integral values and H_s assumes the familiar values $2/(2s+1)$.

A series of values of $H_s(\mu_0)$ has been computed as shown in Table III and by the curves of Fig. 4. The curves have been multiplied by the factors $s+\frac{1}{2}$ in order to make them asymptotic to unity as $1+\mu_0\to0$.

The ratio I_s/H_s is given in Table IV and Fig. 5 rather than I_s itself because it is the ratio which is generally required and because this ratio tends to a finite limit at $\mu=1$, whereas I_s itself vanishes there. Both Fig. 4 and Fig. 5 include information obtained from the integral and half-integral Legendre functions in addition to that tabulated.

VI. POTENTIAL DISTRIBUTIONS INVOLVING CONDUCTING CONES

The potential distribution within a spherical condenser having a conical supporting post as illustrated in Fig. 1e will be determined as an example of the use of the quantities which have been developed.

We shall take the inner sphere of radius a and its supporting cone of exterior half-angle θ_0 to be grounded, while the outer sphere of radius b is held at a potential V_0. To make the potential vanish at the inner sphere, we write Eq. (1) in the form

$$V=\sum_s A_s P_{\nu_s}(\mu)[(r/a)^{\nu_s}-(a/r)^{\nu_s+1}].$$

To evaluate the constants A_s, we set $r=b$ where $V=V_0$, multiply both sides by $P_{\nu_t}(\mu)$, and integrate. All terms except one vanish from the sum by virtue of the

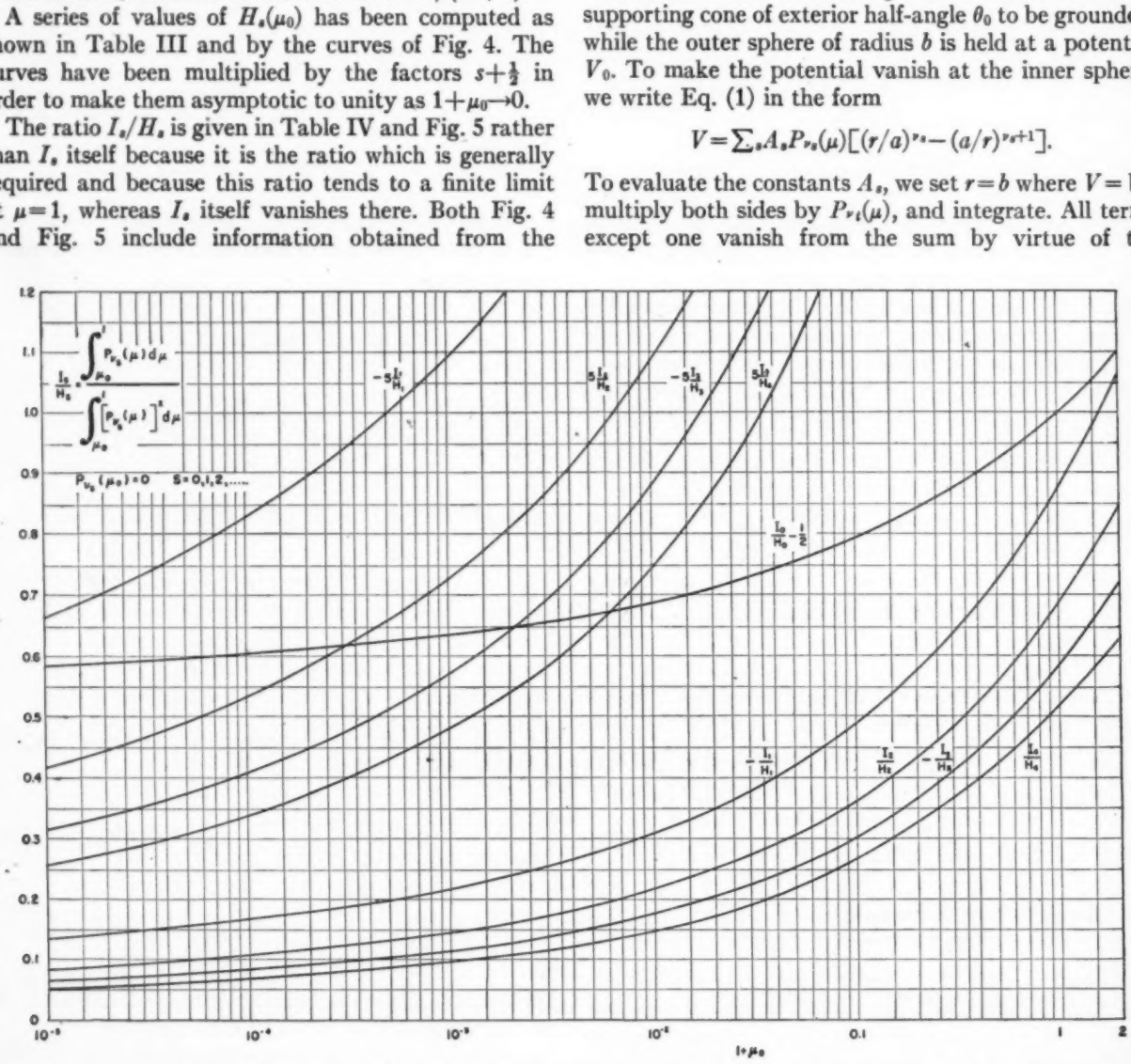


FIG. 5. Integrals of Legendre functions divided by integrals of their squares from $\mu=1$ to the $(s+1)$th zero located at $\mu=\mu_0$.

orthogonality property expressed by Eq. (6), thus determining the constants. We find

$$V=V_0\sum_{s=0}^{\infty}\frac{I_s(\mu_0)}{H_s(\mu_0)}\frac{(r/a)^{\nu_s}-(a/r)^{\nu_s+1}}{(b/a)^{\nu_s}-(a/b)^{\nu_s+1}}P_{\nu_s}(\mu). \qquad (11)$$

In order to obtain the potential distribution for a given cone angle θ_0, we read off successive values of ν_s and I_s/H_s from the curves, giving the required series. Figure 6 gives the factor γ by which the field strength at the point P $(r=a,\theta=0)$ is reduced by the presence of the supporting cone for various sphere radii and cone angles.

The potential distributions about the electrodes shown in Figs. 1a, b, c are easily obtained as limiting cases of Eq. (11). Thus the field strength along the axis of a single cone is proportional to r^{ν_0-1}, ν_0 being the smallest index which causes the Legendre function to vanish on the cone. For mechanical reasons, it is difficult to construct cones whose included half angles are less than 0.05 radian corresponding to $1+\mu_0=10^{-3}$. For such a charged cone, the field strength on the axis is proportional to $r^{-0.87}$ as compared with a variation of r^{-1} for the limiting case of extreme sharpness.[8]

The field distribution about the ring and cone geometry of Fig. 1d may be obtained by the method illustrated by Smythe,[1] Section 4.07. The potential due to a ring located at $r=a$, $\theta=\alpha$ and carrying a charge Q in the presence of the earthed cone $\theta=\beta$ is easily found to be

$$V=\sum_{s=0}^{\infty}A_s(r/a)^{\nu_s}P_{\nu_s}(\cos\theta) \quad r<a$$

$$=\sum_{s=0}^{\infty}A_s(a/r)^{\nu_s+1}P_{\nu_s}(\cos\theta) \quad r>a$$

where

$$A_s=\frac{Q}{a(\nu_s+\frac{1}{2})}\frac{P_{\nu_s}(\cos\alpha)}{H_s(\cos\beta)}.$$

By letting α approach zero, we obtain the field due to a point charge on the axis of a conducting cone.

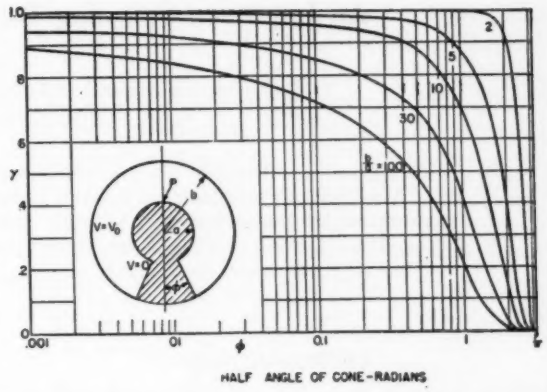


FIG. 6. The factor γ by which the field strength is reduced in a spherical condenser at the point P due to the presence of a conical supporting post. P is located on the inner sphere opposite the post.

ACKNOWLEDGMENT

The foregoing work grew out of an investigation of the field distribution within a photo-cell of the spherical condenser type. The problem was suggested by Dr. L. Apker of this Laboratory without whose frequent encouragement the investigation would not have been carried out.

[8] This distribution may be compared with the potential surrounding a charged paraboloid of revolution as given by $V=\log(Z^2+\rho^2)^{\frac{1}{2}}+Z$ where ρ is the radius in cylindrical coordinates. Along the axis, $V=\log 2Z$ so the field strength varies inversely with distance along the axis as in the case of the sharp cone. This implies that the field about an extremely thin semi-infinite conductor is nearly independent of any variations in cross section along its length. The extent to which this is true may be judged by comparing the potential distribution surrounding cones of varying degrees of sharpness.

An Improved Spinning Top Homogeneous Spray Apparatus

K. R. May
Microbiological Research Department Ministry of Supply, Wiltshire, England
(Received February 3, 1949)

An apparatus is described which produces homogeneous mists or clouds of solid particles of any desired size. As in Walton's original model liquid is sprayed by a Beams high speed air-driven top, but an improved performance has been achieved by utilizing a property of high speed air films so that automatic extraction of unwanted satellite droplets, better running characteristics over a wider range, simplified construction, and low air consumption are obtained.

INTRODUCTION

AN apparatus has recently been described by Walton and Prewett[1] in which the air-driven, air-supported high speed top of Beams[2] is adapted so that liquid fed on to the flat rotating surface is flung off as a very nearly homogeneous mist, particle size being controlled by the speed of the rotor. Walton and Prewett point out that numbers of smaller satellite droplets are formed in addition to the main droplets and the difference in projection distance from the rotor of the main and satellite droplets can be utilized to allow only the former to escape from the apparatus. The present paper describes a development of this apparatus in which improvements have been made as follows:

(a) The apparatus comprises fewer parts than the original,
(b) it is much lighter to give stable running by lessening inherent mounting difficulties,
(c) a wider range of droplets can be produced,
(d) there is less turbulence in the air surrounding the spinning surface so that the apparatus runs cleaner,
(e) an interesting property of high speed films of air moving over continuous solid surfaces enables the working air to be led away and discharged without the need for external suction and at the same time the unwanted "satellite" droplets are entrained in the working air for subsequent disposal,
(f) compressed air requirements are very small so that a small bench type of pressure pump will operate the spray.
(g) Apart from the spraying capabilities of the apparatus the modifications made to the rotor and stator appear to give general running characteristics in advance of anything previously attained with the air-driven spinning top system.

The remarkable steadiness and constancy of rotation of the air-driven top and the ease with which the necessary fairly high rotational speeds can be obtained and controlled give the system outstanding advantages over any other form of high speed drive.

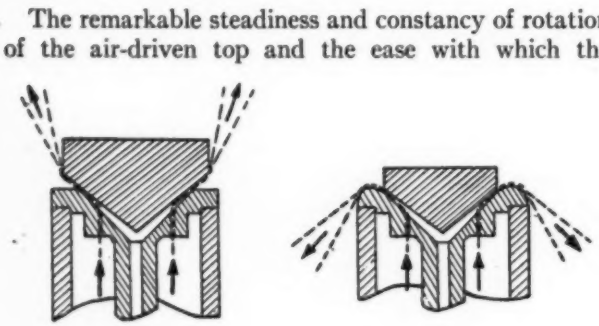


FIG. 1. Flow of working air past large and small rotors.

[1] W. H. Walton and W. C. Prewett, Proc. Phys. Soc. (in the press).
[2] J. W. Beams, J. App. Phys. 8, 795 (1937).

GENERAL DESIGN CONSIDERATIONS

In the Walton and Prewett apparatus the stator and rotor dimensions closely follow the proportions given by Beams, here illustrated in Fig. 1a, where it will be seen that the rotor is greater in diameter than the stator cone. In such a system the air, after impingement on the flutes of the rotor, is found to follow the path indicated by the dotted lines i.e., the film of air tends to cling to the rotor surface and is discharged upward. In use it was found difficult to exceed 2000 r.p.s. with the Fig. 1a arrangement (the stator diameter being 1 inch). Only massive rotors were stable at this speed and a further increase in air pressure usually resulted in the rotor becoming unstable and jumping the stator. In the light of the development described below it now appears that the instability above certain critical pressures may be due to the upward flow of spent working air around the rotor, the surface drag of which produces an upward force on the rotor inadequately compensated by the Bernoulli suction effect within the stator which normally holds down the rotor. It is possible also that the turbulence of the diverging annulus of spent air induces vibration in the rotor. A further serious disadvantage of this upward stream of air as far as the production of homogeneous droplets is concerned is that, even with the close fitting guard ring around the rotor in the Walton and Prewett apparatus, spent air escapes upwards producing turbulent conditions in the annular space from which the generated droplets are flung. This turbulence increases the chance of collisions between droplets thus spoiling homogeneity, and also necessitates the application of a powerful suction to the apparatus considerably in excess of the rate of supply of compressed air, in order to overcome the turbulent effects and ensure that satellites do not escape.

The elimination of these difficulties became possible after the chance use of a rotor of *smaller* diameter than the stator, this system being the essential feature of the design described in this paper. When the small rotor was tried it was found to be completely stable up to the highest pressure (about 80 lb./sq. in.) then available and ran up to 4000 r.p.s. At the same time the

unexpected discovery was made that instead of the spent air discharging upward it now left the stator in a downward direction. The outer lip of the stator was then rounded as shown in Fig. 1b to encourage this effect and the spent air flowed away as indicated by the dotted lines. Comparison of Figs. 1a and 1b shows that the difference in discharge direction follows from the reluctance of the thin film of high speed air escaping between the rotor and stator to leave a continuous solid surface, in both cases the air follows the last continuous surface and tends to flow around the corners of that surface rather than continue straight on into free air. This supersonic flow-around-a-corner effect appears to be similar to that discussed theoretically by Prandtl and Meyer[3] as an irrotational motion problem and the practical results of the effect in this case are of importance for the following reasons:

(a) A correctly made rotor, whether of heavy or light metal, runs with complete stability at the highest available pressures so that much higher rotational speeds can be obtained than with the Fig. 1a type of system;

(b) the upper surface of the rotor is free to perform its spraying function in comparatively undisturbed air, thus minimizing undesirable turbulence effects.

(c) The high velocity film of air emerging from the stator may be usefully employed (vide infra) to suck away the unwanted satellite droplets by an ejector effect and entrain them in the discharged working air which may be ducted away.

DESCRIPTION OF APPARATUS

A scale diagram of the apparatus incorporating the above principle is shown in Fig. 2. The liquid to be sprayed is introduced through the tube (*A*) on the lower end of which is a taper to fit the shortened hypodermic needle (*C*). The clearance between the tip of the needle and the flat upper surface of the rotor (*F*) may be adjusted by means of the bush (*B*) into which the tube (*A*) is a good sliding fit so that it may be withdrawn quickly when required. The cap pieces (*D*) and (*E*) are machined from the transparent plastic "Perspex" both for lightness and to allow the jet adjustment to be observed, though (*E*) may be made of light alloy if desired. The bearing surface between (*D*) and (*E*) is greased and (*D*) is movable relative to (*E*) so that after adjusting the needle tip just to clear the rotor surface, it can be precisely centered over the rotor, an important point for correct spraying. The 45° cone in (*D*) forms an observation window. The under surface of (*E*) has a slight downward and upward slope so that any liquid on it tends to run away down the three slender support pillars on to which it is a push fit. The external diameter of (*E*) is 2 inches.

The rotor rests on the inner assembly which consists of the brass stator (*G*), the aluminium alloy cylinder (*H*) and the bottom plate and compressed air inlet tube, the parts being clamped by the nut at the bottom. The outer assembly consists of the stainless steel dome-shaped piece (*J*) and the aluminium alloy base (*K*) to which are welded the three $\frac{3}{4}$-inch O.D. support legs with air exit ports. The upper surface of the dome (*J*) has a slight inward slope (2°) to assist liquid drainage and the diameter of the hole encircling the rotor allows only 1/1000 in. clearance for the under shoulder of the rotor on withdrawal. The internal surface of (*J*) is smooth and polished to assist the escape of the high velocity spent air.

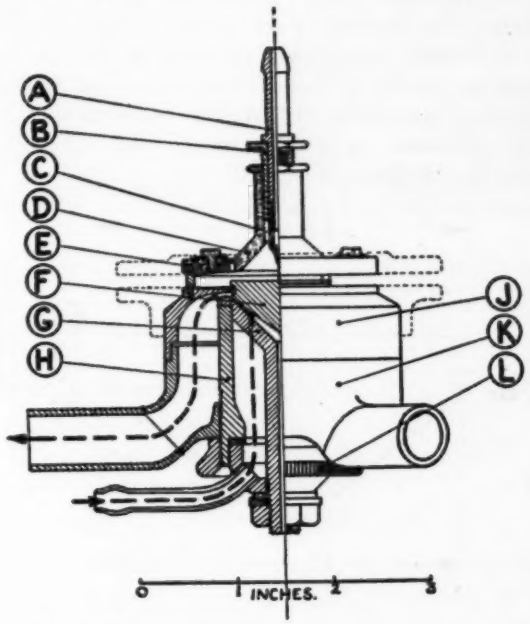


FIG. 2. Diagram of homogeneous spray apparatus.

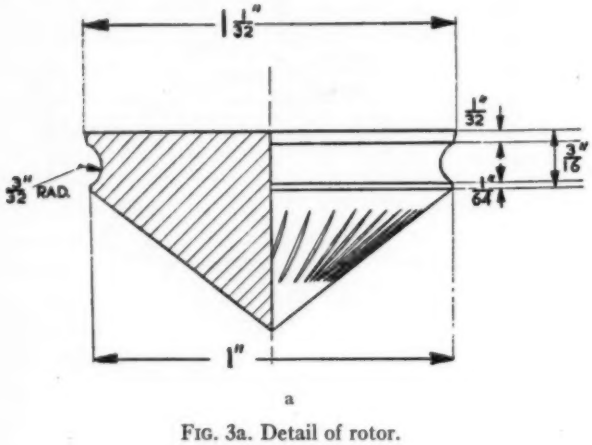
The rotor and stator are shown in greater detail in Fig. 3. The rotor is made of steel and the upper surface and edge should be true and sharp and should be protected from mechanical injury. The angle of the rotor cone is 104°. The number of flutes is not critical but forty are used in this case. Deep cuts should be avoided as there is good evidence that they give rise to instability. Straight shallow cuts should be made such as are given by a 3-inch diam. 45° milling cutter, the resulting flutes being about $\frac{1}{4}$-inch long and inclined to the radii at about 10 degrees.

The stator cone is 91° and the vent hole through the stem is made with a number 32 drill. The four jets are made with a number 65 drill and great care and precision are required to ensure that the angle of emergence of the jets into the cone is correct. It is advisable to make a jig for drilling the jets so that reproducible stators can be made. In Fig. 3 the lines *Y* and *Y'* are jet axes. The distance of 0.52 inch between the lines *YY* (which is actually the separation of the planes through the jet axes parallel to the stator axis) is twice the amount of off-set of the jet drill from the stator axis. Where a jig is not used, the drilling procedure is to set up the drill at 55° to the stator axis as shown by *Y'*

[3] Ewald, Pöschl, and Prandtl, *The Physics of Solids and Fluids* (Blackie & Son, Ltd., London and Glasgow, 1936), second edition, p. 367.

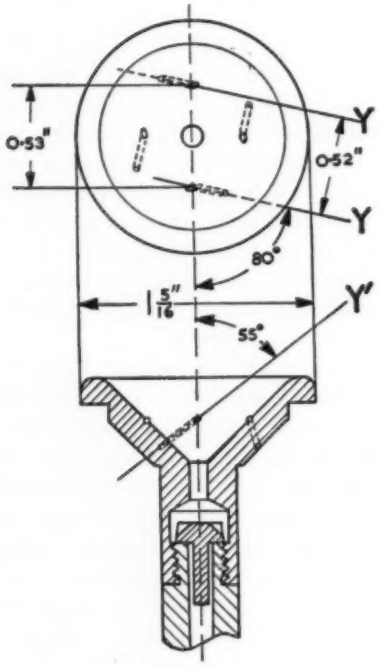
and off-set the drill 0.26 inch, drilling from the underside of the stator, the stator cone having been left unfinished. Having drilled to an adequate depth the stator is returned to the lathe and the cone turned out until the centers of opposite pairs of jets on emergence are 0.53 inch apart when it should be found that Y makes the angle of 80° as shown. The turning over of the rim of the stator is commenced from the point where the diameter is 1 inch. The mushroom valve shown in the stator stem needs to be of minimal weight and is therefore made of plastic material such as Bakelite. The function of this valve is explained later.

Lightweight components are used in the apparatus wherever possible to minimize inherent mounting difficulties as it was found with Walton's original spray that each additional fitment increased the difficulty of mounting the spray to run smoothly.



a

FIG. 3a. Detail of rotor.



b

FIG. 3b. Detail of stator.

The apparatus should be mounted with its legs resting in vee-notches cut in $1\frac{1}{4}$-inch high rubber bungs. The bungs should be as closely spaced as the legs allow and may be stuck to a sheet rubber base. It is important that this should rest freely on a rigid, level surface as the spray is unlikely to run satisfactorily if in any way clamped down or restrained. If connections are made to the exhaust ports to lead off the used air, only very thin walled rubber tubing exercising negligible mechanical restraint should be used. Similarly the connections to the pressure and liquid inlets should be as flexible as possible.

MODE OF OPERATION

When the apparatus is running, the working air follows the path indicated by the heavy broken lines in Fig. 2. In the gap between the dome (J) and the rim of the stator (G) the high velocity stream of air gives rise to an ejector effect and sucks air into the annular gap between (E) and (J) and then down through the gap between the rotor and (J). This inward flow of air draws back the fine, unwanted satellite droplets and by virtue of the greater projection distance of the larger homogeneous droplets allows only the latter to escape from the apparatus. The satellites are therefore automatically separated from the main droplets and are entrained in the spent working air. The majority are impacted on the internal surfaces of the apparatus and any not thus trapped are carried out of the exhaust ports by the air stream.

The function of the mushroom valve in the stator stem shown in Fig. 3b (but not in Fig. 2) is to improve the starting, slow running and stopping characteristics of the rotor. Conditions in the stator cone are such that at working air pressures below 4 lb./sq. in. a positive pressure exists at the apex. Without the valve, therefore, part of the working air flows down the stem instead of supporting the rotor which thus tends to come into intermittent unstable contact with the stator. The valve ensures that all the working air fulfills its correct function in supporting and turning the rotor down to the lowest pressure which will just lift the rotor. The rotor will actually run smoothly at an air pressure as low as 0.12 lb./sq. in. in (6 mm of mercury) with the valve in place, compared with a minimum pressure of 2–3 lb./sq. in. without the valve. At about 4 lb./sq. in. the pressure at the top of the stator vent is equal to atmospheric and as the working pressure is increased, pressure at the top of the vent decreases, the valve lifts, and the necessary upward stabilizing flow for high speed running[2] is induced.

It may be pointed out here that since the satellite droplets are only about one-quarter of the diameter of the "main" droplets, and at moderate rates of liquid feed there are about four satellites per main drop, less than 10 percent by volume of the liquid dispersed appears as satellites. There may be cases in which this small percentage is of no importance and the elaboration of

the automatic satellite separation device described above would then hardly be worth while. It would then be sufficient to fit the system shown in Fig. 1a or 1b with an adjustable support for the liquid feed needle and with the stator stem valve. When fine, the cloud of droplets would then be entrained with the spent working air moving as shown in Fig. 1.

PERFORMANCE DETAILS

Typical performance curves relating working air pressure to main drop size and air consumption over the range of the apparatus are shown in Fig. 4. Most liquids give curves close to that shown for "oils." The water curve is displaced upward by the high surface tension. A wetting agent in aqueous sprays lowers the curve considerably. The very modest requirements for compressed air in the $>20\mu$ range will be noted. Rotational speeds are not given here as there is no necessity to measure them in normal operation when a pressure gauge and sensitive control valve are all that is required. The relation between disk speed and drop size is fully discussed by Walton and need not be repeated here.

When the spray is running, the projected droplets on being arrested by air resistance appear as a well-defined ring-in-space around the apparatus (Fig. 5).

The radius of this ring varies a little according to the liquid sprayed but is directly proportional to the droplet size (Fig. 6). This phenomenon could be deduced by

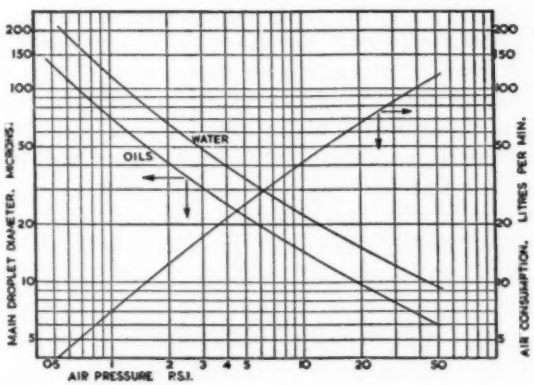


FIG. 4. Performance curves of homogeneous spray apparatus showing relationship between drop size and working air pressure and consumption.

assuming that the projected droplets obey Stokes' law. Consideration of the Reynolds number of the droplets at the moment of projection, however, shows that Re is outside the Stokes' law range. Thus if in Walton's equation for the spinning disk sprayer (1)

$$d=k/\omega(T/D\rho)^{\frac{1}{2}},$$

where d = droplet diameter, ω = angular velocity of disk, T = surface tension of liquid sprayed, D = disk diameter and ρ = liquid density, we put 4.5 for k which is the mean value observed by the author and which is in

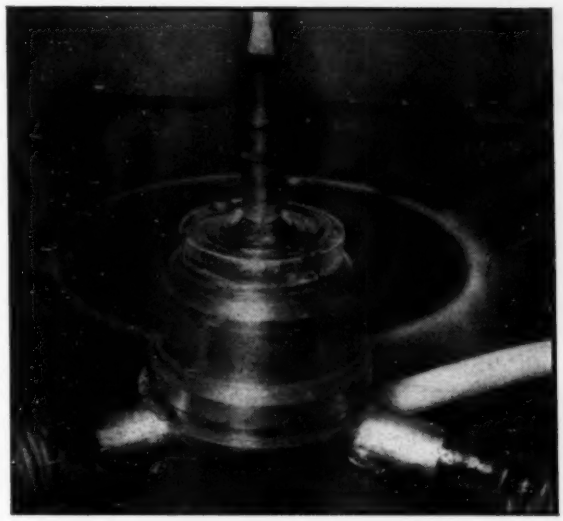


FIG. 5. Appearance of homogeneous spray of 40μ-droplets in powerful light beam (10 sec. time exposure). The ring of droplets was illuminated by a thin horizontal sheet of light and the droplets falling vertically from the ring are therefore not visible.

agreement with Walton and Prewett's figure for their air-driven top we have

$$d=4.5/\omega(T/D\rho)^{\frac{1}{2}},$$

whence

$$vd/\nu=Re=\omega Dd/2\nu=2.25/\nu(TD/\rho)^{\frac{1}{2}}$$

where ν = the kinematic viscosity of the surrounding gas. Putting $\nu=0.14$ for air, $T=25dy/\text{cm}$ say, $D=2.6$ cm and $\rho=1$ g/cm^3 we have $Re\approx 130$ for the majority of non-aqueous liquids.

Stokes' law holds when $Re<1$ which only obtains when the droplets are near the end of their horizontal travel. It should be noted that the value of Re for the system is independent of drop size and disk speed. If the distance of projection x is a function of the particle diameter d, velocity of projection v, viscosity of the air η, and particle density ρ, straight-forward dimensional analysis shows that $x/d=\text{const}$. One should actually measure x along the tangent to the disk since

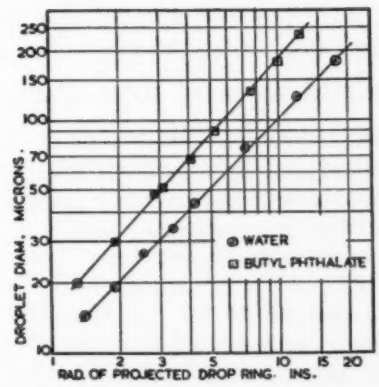


FIG. 6. Relationship between droplet diameter and the distance of projection from the center of the apparatus.

the drops are flung off tangentially but there is no serious error in regarding this distance as the radius of the ring of projected droplets except at the smallest sizes. Although the relationship $x/d=$ const. holds very closely in practice, the mechanics of the droplet's travel

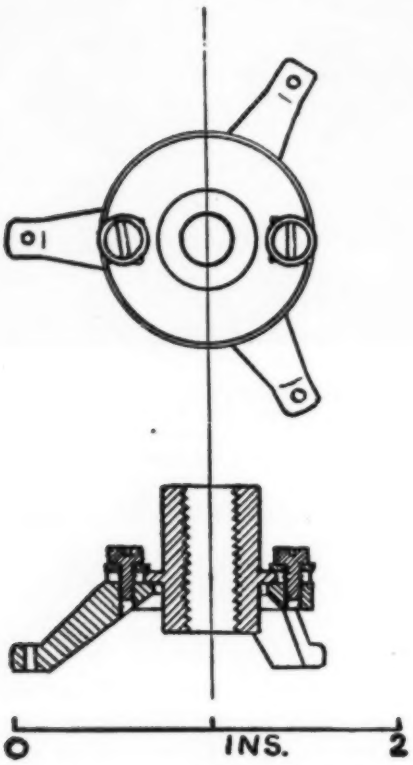


FIG. 7. Spider, supporting liquid feed tube when spraying small droplets.

must, in fact, be complex since flow conditions change from turbulent to viscous and the projection velocity is such that appreciable flattening of the drops from aerodynamic forces is to be expected.

Owing to entrainment of air the radius of the ring of droplets increases slightly with the rate of feed, nevertheless the ring radius provides an invaluable guide to the size of droplet being generated. Thus for water (Fig. 6) it happens that the drop diameter in microns is ten times the ring radius in inches.

RANGE OF DROPLET SIZES AVAILABLE

The upper limit of available drop size is fixed by the minimum speed at which the rotor will run steadily. Provided that the air supply is steady and the mounting satisfactory the rotor will run steadily at less than 1 lb./sq. in. producing drops $>150\mu$. The satellites of such large droplets are themselves too large to be separated automatically by the apparatus, but the relatively high terminal velocities of drops $>100\mu$ enables the separation to be easily effected when required by allowing the main droplets to fall through an annular slot of the appropriate radius in a disk concentric with the generating disk and about an inch lower.

When working in the droplet range of roughly 45–100μ it is necessary in order to ensure automatic satellite separation, to increase the diameter of the exit annulus above the 2 inches shown in Fig. 2 by fitting the plastic rings shown by dotted lines. These merely rest in place and are $3\frac{1}{2}$ inches in diameter so that they can cover the projected distance of the satellites of the main drops in question. (Satellites are about one-fourth of the main drop size and therefore have one-fourth of the projection distance.)

A further point when spraying large droplets is that since the air pressure is low, the suction developed is small and it may be found desirable to apply suction to the exhaust ports to ensure complete satellite removal.

At the lower end of the available range, droplets $<15\mu$ fail to escape from the Fig. 2 arrangement owing to their small projected distance and the high suction field. The cap pieces (D) and (E) are therefore abandoned and the liquid feed tube is supported by a three-legged spider (Fig. 7). With this, the suction field immediately surrounding the rotor retains the satellites while a cross draught will carry off the main droplets. Mists down to 6μ have been produced in this way. In general, however, it is more convenient to use the Fig. 2 arrangement and spray 20μ–40μ droplets of solutions or stable suspensions of controlled concentration so that after evaporation of the volatile solvent droplets or solid particles of the required size remain. Any size of small droplet or particle can be produced in this way. A vertical wind-tunnel with a controlled, non-turbulent updraught of air is advisable to ensure adequate time for evaporation and to prevent aggregation of droplets. It must be noted that highly volatile solvents cannot be used, as with such there is time for the evaporation to take place from the thin, relatively slowly moving but well ventilated film of liquid on the rotor surface so that the concentration can change greatly by the time the rotor rim is reached.

DEGREE OF HOMOGENEITY

As the radius of the ring of droplets appearing in space around the apparatus is directly proportional to the droplet diameter the sharpness of the ring (best seen in a powerful light beam) gives a visual measure of the homogeneity of the cloud (see Fig. 5). A diffuse band instead of a sharp ring indicates that either

(a) the rotor surface is unevenly wetted by the sprayed liquid,
(b) the needle is out of adjustment,
(c) the rotor surface and edge are damaged, or
(d) the rate of liquid feed is too high.

To study the size variation present with the instrument in good adjustment the spray was impacted directly on a glass slide and by methods described elsewhere by the writer[4] traverses were made across the

[4] K. R. May, J. Sci. Inst. 22, 187 (1945).

slide with the microscope, sizing each drop crossing the eyepiece scale, the drops being sized to the nearest micron with an estimated accuracy of slightly better than $\pm\frac{1}{2}\mu$. Both the so-called "absolute" and "magnesium oxide" methods of obtaining permanent droplet samples were employed. The two methods gave results in close agreement but the latter was more commonly used as it is simpler and not subject to possible errors due to droplets coalescing on impact. Between 150–200 droplets were sized for each sample and a range of 30–80μ was investigated. Typical histograms thus obtained are shown in Fig. 8. While the attainable accuracy of measurement of individual drops was insufficient in relation to the small size scatter to enable accurate size frequency distribution curves to be drawn, the histograms obtained indicated that the droplets are distributed symmetrically about the mean size in a nearly normal distribution curve of a small standard deviation. Over the best working range of the instrument and for a given liquid the standard deviation, expressed as a percentage of the mean size, is independent of the absolute value of the mean i.e., the coefficient of variation is constant. This statement should be qualified by saying that there is some evidence to show that the size scatter may increase as the droplet size diminishes to a figure comparable to the roughness of the rotor edge. Thus a deliberately roughened rotor will spray large droplets as efficiently as a smooth one but when fine droplets are sprayed the size scatter from the rough rotor is greater than from the smooth rotor. The more perfect the finish of the rotor edge, therefore, the better the fine droplet performance.

Normally when spraying fine liquids which readily wet the rotor (oils, organic solvents, etc.) a value of about 2.3 percent of the mean drop size is obtainable for the standard deviation, with 90 percent of all droplets falling within a band of 5 percent of the mean in width and the minimum observed drop 6 percent smaller than the mean. These figures can be improved by very accurate machining of the upper surface and edge of the rotor. Water does not readily wet the metal surface of the rotor and is not easily persuaded to spread over the surface as the thin uniform film which is essential for correct spraying. For this reason water or aqueous suspensions or solutions are less favorable for spraying than other liquids. With the rotor surface treated as described later, a typical water spray showed a standard deviation of 5 percent of the mean, a 90 percent band width of 12 percent of the mean and a minimum drop 9 percent smaller than the mean.

When the "main," nearly homogeneous droplets are smaller than about 150μ, they are always accompanied by a small number of larger droplets. It is possible that the proportion of "larger" droplets increases very slowly as the main size diminishes but over the best working range of the apparatus their number varies between one and two percent of the main droplets. Their mean volume is twice that of the main droplets ($D=2^{\frac{1}{3}}d$, where D and d are the drop diameters) with about the same amount of size scatter as the main drops. On a mass basis they comprise two to five percent of the total sprayed liquid emerging from the apparatus. In addition a very small number of $3^{\frac{1}{3}}d$ droplets (0.5 percent) is sometimes found.

The origin of these larger droplets is obscure as no relationship has been found between their occurrence and the physical constants of the sprayed liquid or the adjustment of the apparatus. They may be removed, if desired, by arranging a hoop around the apparatus of diameter slightly greater than the projected distance of the main droplets. The larger droplets, whose projection distance exceeds that of the main droplets, are trapped by impaction on the hoop and an up or down draught will carry away the main droplets. Clearly a different sized ring is required for each change in main drop diameter.

OPERATIONAL NOTES

Needle Adjustment

It has already been mentioned that the liquid feed needle must be precisely adjusted over the center of the rotor. Too great a clearance or an off-centering of the needle allows partial drop formation at the needle tip giving intermittent flow over the rotor. For most liquids a running clearance of about $\frac{1}{64}$ in. is satisfactory.

Rate of Liquid Feed

At excessive rates of liquid feed the rotor tends to become unsteady and the large number of outgoing droplets begin to entrain satellites. These effects do not occur below 1 cc/min. which represents the maximum spraying rate. The rate of feed can be controlled by the head of liquid and the bore and length of hypodermic needle used. Connection between the liquid reservoir (e.g., a burette with side tube) and the spray should be a length of thin-walled fine bore rubber tubing so that no mechanical restraint is exerted on the spray and air locks are prevented.

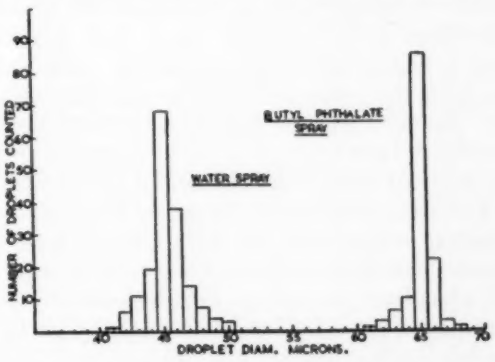
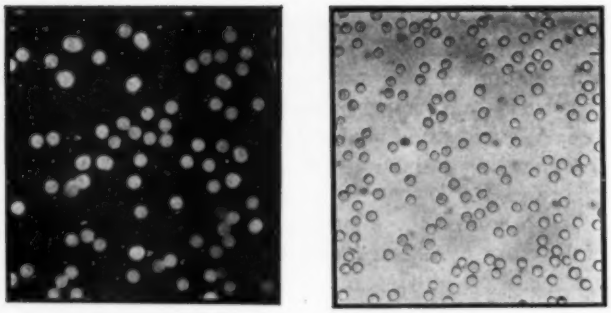


FIG. 8. Typical histograms showing distribution of "main" droplets.



a b

FIG. 9. Photo-micrographs illustrating homogeneous dispersions of (a) an organic liquid, sampled by impact on an MgO coated slide (b) dried clusters of micro-organisms after spraying from aqueous suspension.

Starting and Stopping the Rotor

The gap between the stator rim and the dome-shaped cap (*G* and *J* in Fig. 2) should be adjusted to between $\frac{1}{32}$ in. and $\frac{1}{16}$ in. to ensure self-starting and correct air in-flow conditions. When the valve in the stator stem is fitted, the rotor will start at a very low pressure and no precautions are needed in turning on the air supply.

When stopping the rotor, it is advisable to turn down the air supply to say 0.5 lb./sq. in., allow the rotor time to coast down to equilibrium speed and then turn off the air. In this way violent unsteadiness with consequent wear and tear is prevented. The purpose of the very small clearance between the lower shoulder of the rotor and the encircling hole in *J* (1/1000 in.) is to prevent the rotor jumping out of its seating before coming to rest.

Spraying Water

Water and water suspensions and solutions are always more difficult to spray satisfactorily than oily or solvent liquids which readily wet the rotor. It is possible to achieve a satisfactory water spray by removing all grease from the rotor surface by scrubbing with wet pumice powder and also adding a wetting agent to the water. A better technique is to apply the process known as "parkerizing" (developed by the Parker Rust Proof Company, Detroit) to the rotor. The fine crystalline coating produced on the surface of the rotor after 15 minutes boiling in the parkerizing solution wets very readily when fresh, and sprays well. A similar crystalline coating may be given to aluminium rotors by 10–15 min. immersion in a hot solution of phosphoric acid followed by a nitric acid rinse. Recent work has shown that ground glass is likely to be the best surface for spraying water. The edge and upper surface of a thin glass circle are ground as accurately as possible with the finest grinding powder in water. The circle is cemented to the rotor surface. When dirty, the glass surface is very easily renovated by a few seconds regrinding on a flat glass plate.

The non-wetting of metal and Perspex by water gives poor drainage from the annulus around the rotor so that the water gathers in pools or large drops, blocking the exit of the sprayed droplets. A wetting surface which drains well may be made by brushing a film of gelatine solution plus a wetting agent over the internal surfaces of the annulus.

Spraying Solid Solutions or Suspensions

Evaporation from the indrawn satellite droplets builds up a deposit in the gap between the stator and dome when solid particles are being produced by spraying solutions or suspensions of solids. The running time before cleaning becomes necessary and is therefore limited to something of the order of 30 minutes depending of course on the concentration used. To avoid build up on the rotor rim strong solutions in volatile solvents should not be used.

APPLICATIONS

The field of application of homogeneous dispersions is very wide particularly as sprays of this type have not hitherto been available. The effect of particle size in insecticidal and other forms of aerosols, the performance of aerosol sampling apparatus, the efficiency of industrial dust filters, the mechanism of infection by airborne micro-organisms, readily suggest themselves as fields in which new information will be obtainable and there are doubtless many others. Fig. 9(a) shows a sample in MgO of an organic liquid as sprayed from the apparatus and Fig. 9(b) shows 13μ spherical clusters of the organism Chr. Prodigiosum sprayed from aqueous suspension as 30μ droplets and allowed to dry in the air before sampling. Each cluster contains about 2000 organisms.

ACKNOWLEDGMENTS

The author wishes to thank Mr. W. H. Walton for reading the manuscript and for helpful criticism and the Chief Scientist, Ministry of Supply, for permission to publish this paper.

Distribution of Temperature and Current in Cylindrical Bodies

CARL F. MUCKENHOUPT
Branch Office, Office of Naval Research, Boston, Massachusetts
(Received February 25, 1949)

The current and temperature distributions in a conductor are affected not only by the magnetic effect of the current, i.e., skin effect, but also by the variation of resistivity with temperature. Consequently the two are interdependent, and since the resistance change amounts to between three and four percent for every 10°C rise in temperature in metals, and is still greater in insulators, its effect can be quite appreciable. A complete solution for the steady state is given herewith.

I. THE SET-UP OF THE DIFFERENTIAL EQUATIONS

WHEN the current and temperature in a cylindrical conductor of radius R and thermal conductivity k have become steady, all heat Q produced per unit of time passes out through the conductor to the surrounding medium. Consequently if we call T the temperature at any distance r from the center, and i and ρ the current density and resistivity, respectively, at the same distance, we have from the figure:

$$-kA(dT/dr)=Q, \tag{1}$$

where A is the area normal to the heat flow, and for unit conductor length equals $2\pi r$.

Now the heat produced per unit volume is $dQ=i_e^2\rho$, where i_e signifies the effective or r.m.s. value of current density. Combining these facts we obtain

$$-2\pi kr(dT/dr)=\int_0^r\int_0^{2\pi} i_e^2\rho r d\theta dr, \tag{2}$$

and due to the radial flow of heat and longitudinal flow of current which will be assumed, the variables T and i_e are independent of θ. So

$$-2\pi kr(dT/dr)=2\pi\int_0^r i_e^2\rho r dr. \tag{3}$$

Differentiating with respect to r and rearranging:

$$(d^2T/dr^2)+(1/r)(dT/dr)+(1/k)i_e^2\rho=0. \tag{4}$$

In considering current distribution we proceed in the well-known manner employed for skin effect alone, and assume that the difference in $i\rho$ drops along two longitudinal filaments at different values of r is due to the effect of the induction e.m.f.

Since there is no radial current the total drop along ab equals that along dc. To compute the e.m.f. induced around the rectangle $abcd$ we use the basic relation

$$H=2I/r, \tag{5}$$

and since the total current inside a radius r is

$$I=\int_0^r 2\pi r i dr, \tag{6}$$

we have

$$H=4\pi/r\int_0^r r i dr. \tag{7}$$

Differentiating and solving for i

$$i=(1/4\pi)(dH/dr)+(H/4\pi r). \tag{8}$$

The total flux ϕ through $abcd$ is $\mu H dr$ and the induced e.m.f. is

$$(d\phi/dt)=\mu dr(dH/dt). \tag{9}$$

Since this accounts for the difference in $i\rho$ drops on the two sides of the rectangle

$$d(i\rho)=\mu dr(dH/dt), \tag{10}$$

which gives

$$(dH/dt)=(1/\mu)[i(d\rho/dr)+\rho(di/dr)]. \tag{11}$$

Now differentiate (8) with respect to t, and (11) with respect to r and eliminate derivatives of H from the expression for (di/dt). We then have

$$di/dt=1/4\pi\mu[i(d^2\rho/dr^2)+2(d\rho/dr)(di/dr)+\rho(d^2i/dr^2)]+(1/4\pi\mu r)[i(d\rho/dr)+\rho(di/dr)]. \tag{12}$$

This is the general equation for current. If we consider a sinusoidal current and neglect the small resistivity variations over a current cycle, i.e., assume the resistivity to depend on the r.m.s. current value, we have, since $di/dt=(d/dt)(i_{\max}\epsilon^{j\omega t})=j\omega i$, $[j=(-1)^{\frac{1}{2}}]$,

$$4\pi\mu j\omega i=\rho(d^2i/dr^2)+[(\rho/r)+2(d\rho/dr)]di/dr+i(d^2\rho/dr^2)+i/r(d\rho/dr). \tag{13}$$

Now call ρ_0 the resistivity at the center of the conductor, T_0 the temperature there, and α the temperature coefficient of resistivity.

$$\rho=\rho_0[1+\alpha(T-T_0)], \tag{14}$$

$$d\rho/dr=\rho_0\alpha(dT/dr). \tag{15}$$

$$d^2\rho/dr^2=\rho_0\alpha(d^2T/dr^2). \tag{16}$$

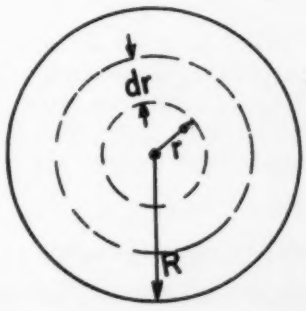


FIG. 1. Cross section of the cylindrical body.

Substituting as required in (13) and in (4) and rearranging we finally get

$$i[(d^2T/dr^2)+(1/r)(dT/dr)]+(1/\alpha-T_0+T)[(d^2i/dr^2)+(1/r)(di/dr)]+2(di/dr)(dT/dr)=Ci. \quad (17)$$

$$(d^2T/dr^2)+(1/r)(dT/dr)+Bi_e^2[(1/\alpha)-T_0+T]=0, \quad (18)$$

where $B=\rho_0\alpha/K$ and $C=j(4\pi\mu\omega/\rho_0\alpha)$.

II. SOLUTION OF THE DIFFERENTIAL EQUATIONS

These equations, (17) and (18) are the simultaneous differential equations of the system. They appear

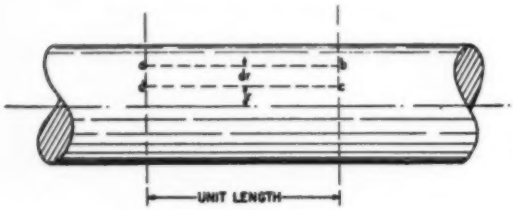


FIG. 2. Longitudinal section of the cylindrical body for current and magnetic field considerations.

rather formidable, and being non-linear do not give much hope of solution in terms of elementary functions. Consequently the usual type of power series expansion will be employed.

Due to the symmetry about the center of the cylinder only even powers need be considered.

Let

$$T=T_0-1/\alpha+a_0+a_2r^2+a_4r^4+a_6r^6+\cdots, \quad (19)$$

$$dT/dr=2a_2r+4a_4r^3+6a_6r^5+8a_8r^7+\cdots, \quad (20)$$

$$d^2T/dr^2=2a_2+3.4a_4r^2+5.6a_6r^4+7.8a_8r^6+\cdots. \quad (21)$$

Let

$$i=I_0+b_2r^2+b_4r^4+b_6r^6+\cdots,+j(c_2r^2+c_4r^4+c_6r^6+\cdots). \quad (22)$$

We shall compute this as the r.m.s. current density since there is a constant ratio between i_{max} and the r.m.s. current, the functional expressions being otherwise identical, i.e., $i_{max}=\sqrt{2}i_e$.

From (22) we get

$$i_e^2=(I_0+b_2r^2+b_4r^4+\cdots)^2+(c_2r^2+c_4r^4+c_6r^6+\cdots)^2. \quad (23)$$

The derivatives of i are computed in the same manner as those of T for substitution in Eqs. (17) and (18). This substitution is laborious but straightforward and need not be given. The relations between coefficients which result however are as follows. From Eq. (18)

$$2^2a_2=-Ba_0I_0^2; \quad (24a)$$

$$4^2a_4=-B(a_2I_0^2+2a_0I_0b_2); \quad (24b)$$

$$6^2a_6=-B(a_4I_0^2+2a_2I_0b_2+2a_0I_0b_4+a_0b_2^2+a_0c_2^2); \quad (24c)$$

$$8^2a_8=-B(a_6I_0^2+2a_4I_0b_2+2a_2I_0b_4+2a_0I_0b_6+2a_0b_2b_4+2a_0c_2c_4+a_2b_2^2+a_2c_2^2), \text{ etc.} \quad (24d)$$

From Eq. (17)

$$CI_0=2^2a_2I_0+2^2a_0(b_2+jc_2); \quad (25a)$$

$$C(b_2+jc_2)=4^2a_4I_0+4^2a_2(b_2+jc_2)+4^2a_0(b_4+jc_4); \quad (25b)$$

$$C(b_4+jc_4)=6^2a_6I_0+6^2a_4(b_2+jc_2)+6^2a_2(b_4+jc_4)+6^2a_0(b_6+jc_6); \quad (25c)$$

$$C(b_6+jc_6)=8^2a_8I_0+8^2a_6(b_2+jc_2)+8^2a_4(b_4+jc_4)+8^2a_2(b_6+jc_6)+8^2a_0(b_8+jc_8), \text{ etc.} \quad (25d)$$

Equation sets (24) and (25) are now used simultaneously together with the known conditions that $a_0=1/\alpha$ and $T=T_0$ when $r=0$ to obtain the coefficients of (19) and (22). The algebraic reductions are omitted, and for convenience of computation the final results are presented as follows. Let $M=(\rho_0\alpha I_0^2)/k$ and $N=(\pi\mu\omega)/\rho_0$, where the units are: α in °C, I_0 in abamperes, ρ_0 in abohms-cm, μ the permeability (numeric), $\omega=2\pi\times$ frequency, and $k=4.18\times10^7$ times the usual tabulated thermal conductivity in g cal./cm³/sec./°C. This k value results from the i and ρ units and the use of $i_e^2\rho$ for dQ.

The magnitude of I_0 must also be considered since I_0 appears in all relations, either directly or in M. Due to skin effect I_0 diminishes very rapidly with increase in N since the current density must be held to reasonable values near the surface where it will be maximum. Let $x=R(4N)^{\frac{1}{2}}$. Then with a copper conductor of 1 cm radius and a frequency near 2200 cycles, $x=10$, and the surface density is about $150I_0$. At 10KC the ratio is some 330,000 to 1! For values of x up to 10 this ratio is easily computed from tables of Bessel's functions, being equal to $(ber^2x+bei^2x)^{\frac{1}{2}}$. For large values of x a good approximation is $\epsilon^x/\sqrt{2}/(2\pi x)^{\frac{1}{2}}$.

III. THE TEMPERATURE SOLUTIONS

Let us now consider the solutions separately. Equation (19) becomes

$$T=T_0-\frac{M}{\alpha}\left[\frac{r^2}{2^2}+\frac{M}{2^4(2!)^2}r^4+\left(\frac{5M^2}{2^6(3!)^2}+\frac{2N^2}{2^2(3!)^2}\right)r^6+\left(\frac{59M^3}{2^8(4!)^2}+\frac{7MN^2}{2^2(4!)^2}\right)r^8+\left(\frac{1263M^4}{2^{10}(5!)^2}+\frac{662M^2N^2}{2^6(5!)^2}+\frac{6N^4}{2^2(5!)^2}\right)r^{10}+\cdots\right]. \quad (26)$$

With metallic conductors the electrical and thermal conductivities are so good that M is small, and its value becomes much more so at high frequencies due to the presence of the I_0^2 factor. As a result the temperature is nearly uniform over the cross section unless very large radii or excessive current densities are used. For copper at 60 cycles, $\alpha=0.0039$, $\rho_0=1720$, $k=3.51$

$\times 10^7$ and $I_0=20$ (200 amperes/cm^2).

$$T=T_0-0.00490r^2-0.000129r^6-\cdots. \quad (27)$$

For iron at 60 cycles, $\alpha=0.003$, $\rho_0=2.5\times10^4$, $k=6.27\times10^6$, $\mu=100$ (chosen arbitrarily). Let I_R the surface density be 20 at radius $R=1$ cm. Then $I_0=4.73$ and

$$T=T_0-0.0223r^2-0.0278r^6-0.000784r^{10}-\cdots. \quad (28)$$

These two examples are typical of good conductors, or of any situation where N is large and M small. For these cases, for all practical purposes (26) becomes

$$T=T_0-\frac{M}{\alpha}\left[\frac{r^2}{2^2}+\frac{2N^2}{2^2(3!)^2}r^6+\frac{6N^4}{2^2(5!)^2}r^{10}+\frac{20N^6}{2^2(7!)^2}r^{14}+\frac{70N^8}{2^2(9!)^2}r^{18}+\cdots\right]. \quad (26a)$$

The ratio of adjacent terms approaches

$$\frac{\mathfrak{F}_n+4}{\mathfrak{F}_n}\sim\frac{4N^2r^4}{[(n+2)/2]^2[(n+4)/2]^2}\to\frac{64N^2r^4}{n^4}\to 0 \text{ as } n\to\infty. \quad (29)$$

So (26a) always converges for finite N and r.

With dielectrics and with poor conductors like carbon M becomes large and N negligible. This means a suppression of skin effect, and the variation of resistivity with temperature can become important.

Consider a carbon rod with $\alpha=-9\times10^{-4}$, $\rho_0=5\times10^6$, $k=1.21\times10^7$, and as before $I_0=20$ and $f=60$ cycles.

$$T=T_0-41.4r^2+0.386r^4-0.00798r^6+\cdots. \quad (30)$$

A very large temperature variation has appeared in spite of the relatively good thermal conductivity. For thin rods this is largely due to the high resistivity, but with thick ones such as used in some resistance stacks and in some heating applications the terms after the second on the right become of importance, and these are dependent on α.

In a true insulator even more pronounced effects appear. Consider an insulating post of steatite. Here $k=10^5$, but further consideration must be given to ρ_0 and I_0 since the effective resistivity is an a.c. one depending on the losses caused by the total in-phase component of the current, and I_0 is this current.

The derivation of the expression for dielectric loss is given in most electrical and electronic handbooks and will not be repeated here. The result is

$$W=(10^{-12}/1.8)fe^2d\tan\delta \text{ watt/cm}^3, \quad (31)$$

where f=cycles/sec., e=r.m.s. volts/cm gradient, d=dielectric constant, δ=dielectric loss angle.

Then, since $W=e^2/\rho$,

$$\rho=(1.8\times10^{12}/fd\tan\delta) \text{ ohm cm}. \quad (32)$$

For most practical dielectrics $\tan\delta$ is substantially independent of impressed potential and of frequency over wide frequency ranges. For low frequencies the value of ρ merges into the d.c. value. At high temperatures, i.e., above about 300°C, the a.c. and d.c. values are also practically identical.

Values of $\tan\delta$ must in general be obtained from tables, i.e., by measurement, and α is then obtained from the computed variations in ρ. An inspection of Eq. (26) shows that α has only a higher order effect at best, so that its variations in turn, since they are not large percent-wise, will have still less influence. Consequently it is reasonable to use average values of α for the temperature range encountered.

It is of interest to note, however, that for steatite type materials a good approximation to a functional analysis can be made, for

$$\tan\delta\doteq ke^{\beta T}, \quad (33)$$

where k is of the order of magnitude of 10^{-3}, and β is about 0.005. T is in °C.

Since

$$\rho_2=\rho_1[1+\alpha(T_2-T_1)], \quad (34)$$

we may introduce (33) into (32) to obtain the ρ's in (34), and then solve (34) for an average α over the range T_1-T_2. This gives

$$\alpha=[1-e^{\beta(T_1-T_2)}]/(T_1-T_2). \quad (35)$$

It is seen at once that this analysis makes α depend only on the temperature difference and not at all on the initial temperature. This of course is not accurate, but the values of α are usable, and further analysis of (35) for $d\alpha/dT$ shows that the changes are small enough to be neglected, as was stated previously. From (35) it is apparent that $\alpha\to-\beta$ as $T_2\to T_1$.

With dielectrics less amenable to such analysis recourse must be had to tabulated values.

Let us take $f=10^6$, $d=6$, and $\tan\delta=3\times10^{-3}$. Then $\rho_0=10^8$ ohm cm$=10^{17}$ abohm cm. Let $\alpha=-0.004$. If we assume an r.m.s. gradient of 10^4 volts/cm for e we get $I_0=10^{-5}$ abampere and $M=-0.4$. Then (26) becomes

$$T=T_0-25r^2+0.625r^4-0.00347r^6+\cdots. \quad (36)$$

Whenever N is negligible (26) becomes

$$T=T_0-\frac{M}{\alpha}\left[\frac{r^2}{2^2}+\frac{M}{2^4(2!)^2}r^4+\frac{5M^2}{2^6(3!)^2}r^6+\frac{59M^3}{2^8(4!)^2}r^8+\frac{1263M^4}{2^{10}(5!)^2}r^{10}+\frac{42713M^5}{2^{12}(6!)^2}r^{12}+\cdots\right]. \quad (26b)$$

A convergence study of this relation is difficult due to the complicated form of the coefficients, but an analysis of equation sets (24) and (25) shows that

$$(n^2/\alpha)A_{2n}=-A_{2(n-1)}A_2-2^2A_{2(n-2)}A_4-3^2A_{2(n-3)}A_6-4^2A_{2(n-4)}A_8-\cdots \quad (37)$$

where A_{2n} is the coefficient of r^{2n} in (26b). From this it can be shown that the ratio of adjacent terms, A_{2n}/

$(A_{2(n-1)})$ is approaching M/ϵ, $(\epsilon=2.718)$ as n gets large. A rigorous proof is not at hand, but it can be definitely said that if $|Mr^2| \leqslant 2.5$ the series will converge. In spite of this restriction we may analyze even the most severe case, i.e., insulators, unless large diameters or excessive currents are present. We can always approach a solution by reducing I_0 to a size consistent with convergence and then studying the effect of a gradual increase.

IV. THE CURRENT SOLUTIONS

The current equation can be treated similarly, and (22) becomes

$$i=I_0\Big[1+\frac{M}{2^2}r^2+\Big(\frac{5M^2}{2^4(2!)^2}-\frac{N^2}{(2!)^2}\Big)r^4+\Big(\frac{59M^3}{2^6(3!)^2}-\frac{3MN^2}{(3!)^2}\Big)r^6+\Big(\frac{1263M^4}{2^8(4!)^2}-\frac{227M^2N^2}{2^4(4!)^2}+\frac{N^4}{(4!)^2}\Big)r^8+\Big(\frac{42713M^5}{2^{10}(5!)^2}-\frac{6076M^3N^2}{2^6(5!)^2}-\frac{173MN^4}{2^2(5!)^2}\Big)r^{10}+\cdots\Big]$$
$$+jI_0\Big[Nr^2+\frac{5MN}{2^2(2!)^2}r^4+\Big(\frac{59M^2N}{2^4(3!)^2}-\frac{N^3}{(3!)^2}\Big)r^6+\Big(\frac{1263M^3N}{2^6(4!)^2}+\frac{MN^3}{(4!)^2}\Big)r^8+\Big(\frac{42713M^4N}{2^8(5!)^2}+\frac{1577M^2N^3}{2^4(5!)^2}+\frac{N^5}{(5!)^2}\Big)r^{10}+\cdots\Big]. \quad (38)$$

In the case of good conductors or wherever M is negligible with respect to N this reduces to the well known series of the Bessel function for skin effect:

$$i=I_0[ber\ r(4N)^{\frac{1}{2}}+j\ bei\ r(4N)^{\frac{1}{2}}]. \quad (38a)$$

Consequently the solutions for copper and for iron will not be written out. In the case of poor conductors or of insulators (38) may be taken as

$$i=I_0\Big[1+\frac{M}{2^2}r^2+\frac{5M^2}{2^4(2!)^2}r^4+\frac{59M^3}{2^6(3!)^2}r^6+\frac{1263M^4}{2^8(4!)^2}r^8+\frac{42{,}713M^5}{2^{10}(5!)^2}r^{10}+\cdots\Big], \quad (38b)$$

with the restriction as previously given that for convergence $|Mr^2| \leqslant 2.5$.

For the carbon rod then we have

$$i=20[1-0.0372r^2+0.00173r^4-0.0000845r^6+\cdots], \quad (39)$$

and for the steatite insulator

$$i=10^{-5}[1-0.1r^2+0.0125r^4-0.00164r^6+\cdots]. \quad (40)$$

V. SUMMARY

1. With materials which are both good electrical and thermal conductors the temperature is nearly uniform across the section and about equal to that at the surface. This is true for all frequencies. The current distribution is very closely given by the usual skin effect equations. Exceptions could occur for the case of very large cylinders and/or heavy current densities, and in this instance the complete Eqs. (26) and (38) will furnish the solution.

2. With materials like carbon, graphite, etc., the thermal conductivity is good but the electrical conductivity relatively poor. These form an intermediate class where skin effect is markedly reduced, and very considerable temperature variation can occur from surface to center. Usually equations (26b) and (38b) apply, but in doubtful cases the general relations (26) and (38) should be tried.

3. When both thermal and electrical conductivities are poor, skin effect vanishes for all practical considerations and the current distribution becomes uniform while the temperature distribution may vary widely. Here the relations (26b) and (38b) are applicable, suitable consideration being given to the convergence conditions.

4. The general equations (26) and (38) show that "intermodulation" effects do indeed result, as one would expect, from the non-linearity of the original relations (17) and (18). These show up as products of various powers of M and N. It is interesting to note that their earliest appearance is with the fourth power of the variable for the imaginary component of the current density, with the sixth power for the real component, and with the eighth power in the temperature distribution. In any case where M and N are both within the same order of magnitude these terms can have very noticeable effects.

5. It is to be noted that (26b) and (38b) apply to direct as well as alternating current.

Precipitation from Solid Solutions of C and N in α-Iron*

CHARLES A. WERT
Institute for the Study of Metals, The University of Chicago, Chicago, Illinois
(Received March 14, 1949)

The formation of precipitates from the solid solutions of C and N in α-iron has been studied by means of the internal friction peak associated with the stress-induced interstitial diffusion of the solute atoms. A transformation law has been found to fit the experimental data. The data are interpreted as showing the following: (1) Fe_3C precipitates in the shape of spheres; (2) an intermediate phase in the precipitation of nitrogen forms in the shape of disks; and (3) within the experimental error there is indicated no continuous nucleation. Using the results of a derivation made by Zener, the mean distance between nuclei is calculated for Fe_3C. For carbon concentrations of about 0.016 wt. percent, this distance is about 800, 2000, and 7000 Angstroms at 27°C, 102°C, and 312°C, respectively. With appropriate experimental conditions, the temperature dependence of the precipitation rate is found to be about the same as the temperature dependence of the diffusion rate of the solute atoms.

INTRODUCTION

MANY physical properties of alloys are greatly affected by precipitation from metallic solid solution. Some of these changes have considerable practical importance. In addition, the process of precipitation and some of the changes accompanying it are of interest in themselves, and many studies have been made in this field. In these investigations of precipitation rates, use has been made of such properties as electrical conductivity, magnetic properties, lattice parameter, density, and microscopic examination. Unfortunately, however, even though the property change may be easily measurable, the determination of precipitation rates from the observed data is in general difficult. The major difficulty which arises is that of knowing the relationship between the property change and the amount of the precipitate. A second difficulty in some investigations is the lack of sensitivity during the early stages of the process or during the entire process if the total amount of precipitate is small.

In the instance of interstitial solid solution in body-centered cubic lattices, it is possible to use for precipitation measurements a property peculiar to this type of material. This property is the internal friction associated with the stress-induced interstitial diffusion of the dissolved atoms. The internal friction of such binary alloys of carbon and nitrogen in α-iron and of carbon, nitrogen and oxygen in tantalum have already been studied in some detail.[1–3] Dijkstra, in particular, has made an extensive semiquantitative investigation of precipitation of nitrogen and carbon in α-iron. Among other things, he found and investigated a metastable precipitate of nitrogen. For an initial concentration of 0.045 wt. percent N and a tempering temperature of 250°C, he observed that the precipitation occurred in two distinct stages. The first stage was a precipitate of unknown chemical composition and the second, the well-known Fe_4N phase. The new phase, which we shall call after his nomenclature N Phase 1, was found to have a higher solubility in α-iron than does Fe_4N. These findings, together with his other observations, appeared promising enough to warrant a more careful study of the general problem of precipitation of N and C in α-iron.

The present paper describes a series of experiments designed to measure quantitatively the rate of precipitation of (a) N Phase 1, (b) Fe_4N and (c) Fe_3C from the supersaturated solid solution. The results are interpreted so that a fairly complete picture of the entire process is presented.

THE EXPERIMENTAL PROCEDURE

The iron used in the present investigation was Puron, a product of Westinghouse. This material has an iron content of 99.95 percent. It is particularly free of metallic impurities. The principal impurity is oxygen, present to the extent of 0.040 percent. The specimens were wires 0.025 in. in diameter and one foot long. These wires formed the suspension of a torsional pendulum whose period was about one second. The wire was suspended in an electrically heated furnace so that it could be brought to the temperature at which the internal friction peak occurs for this frequency. The internal friction of the system was measured directly by setting the pendulum into free vibration and observing the decrease in amplitude over a number of oscillations. This method and apparatus of this type are described in detail by Kê.[4]

The specimens were initially cleaned of residual nitrogen and carbon by a wet-hydrogen treatment at 720°C. This ensured freedom from undesired precipitates. After this treatment, solid solutions of C or N in the iron were formed. Carbon was added by heating the wire at 720°C for a number of hours in an atmosphere of dry hydrogen and *n*-heptane. Quenching the specimen to room temperature retained the carbon

* This research was supported by the ONR Contract No. N-6Ori-20-IV.
[1] J. L. Snoek, Physica **8**, 711 (1941).
[2] L. J. Dijkstra, communicated to A.I.M.E.
[3] T. S. Kê, "Metals technology," Tech. Pub. 2370, A.I.M.E., (June 1948); Phys. Rev. **74**, 9 (1948); Phys. Rev. **74**, 914 (1948).
[4] T. S. Kê, Phys. Rev. **71**, 533 (1947).

in solid solution in a homogeneous supersaturated state. Similar supersaturated solutions of nitrogen were prepared by treating the specimens at 590°C in an atmosphere of NH_3 and dry hydrogen before quenching.

The determination at a given temperature of the rate of precipitation of second phases from these supersaturated solid solutions consisted in measuring the height of the internal friction peak as a function of tempering time. For measurements at a given temperature, the following technique was used: A specimen with the desired concentration of C or N was heated to 720°C or 590°C, respectively, in a hydrogen atmosphere. After one minute, the specimen was quenched in cold water. The height of the internal friction peak was then determined. At this point the specimen was tempered in an auxiliary furnace for a definite time and quenched to room temperature by an air blast. The internal friction peak was again measured. Continued repetition of this tempering process extended the tempering time and ultimately gave a plot of internal friction as a function of total time of tempering. Computation from this curve of the amount of precipitate as a function of the tempering time depended on the following facts: (1) Only those atoms of C or N which remain in solid solution contribute to the internal friction and (2) the height of the peak is directly proportional to the amount of C or N remaining in solid solution.[5] Hence, at any time the amount of precipitate (i.e., the amount of solute which has been removed from solid solution) is proportional to the difference between the initial height (when $t=0$) and the height of the peak at that time. This is illustrated by typical tempering and precipitation curves shown in Fig. 1.

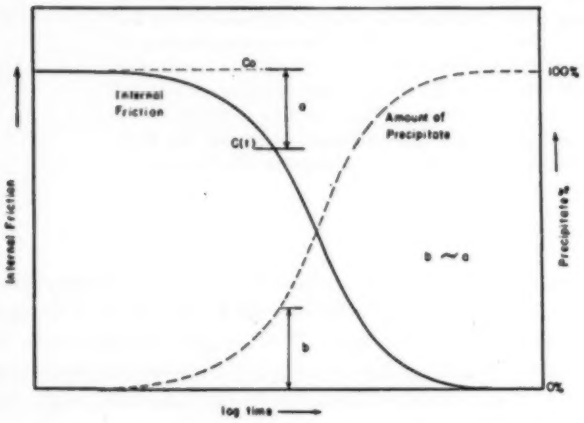


FIG. 1. The amount of precipitate formed at any time is proportional to $C_0-C(t)$.

PRECIPITATION MEASUREMENTS

The rate of precipitation of carbon and nitrogen depends on a number of factors. Among these are temperature, state of cold-work degree of supersaturation, and the presence of impurities. The effect of impurities was not investigated in the present work. Beyond annealing the iron enough to recrystallize it after it had been swagged and drawn from $\frac{3}{8}$ in. to 0.025 in. in

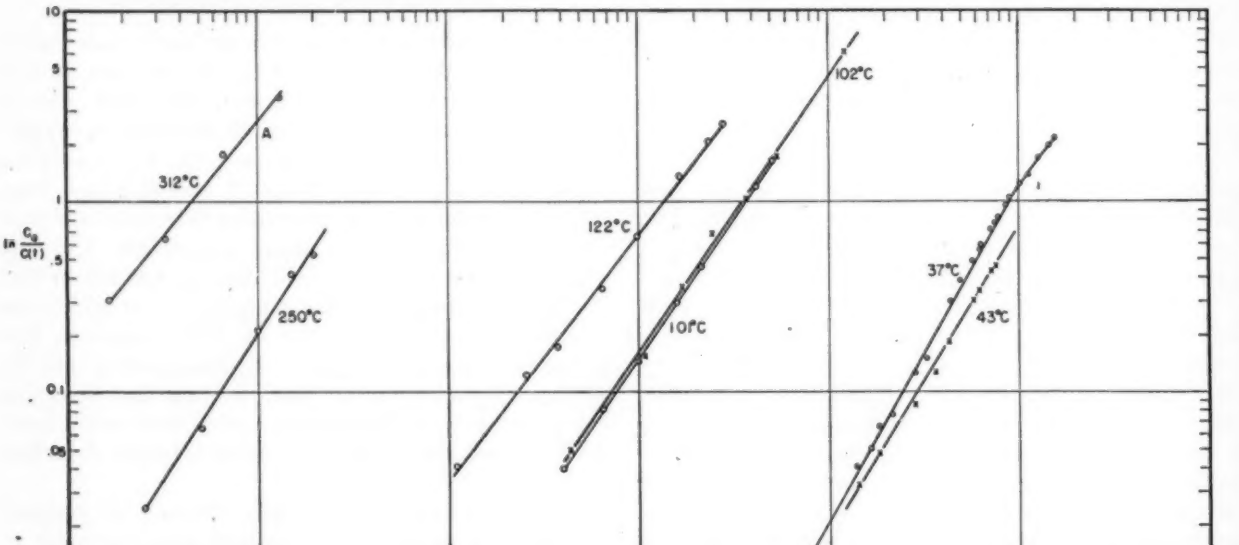


FIG. 2. Precipitation of carbon. The initial concentration of carbon in solid solution was about 0.016 percent by weight for all curves except curve A, for which it was about 0.019 wt. percent.

[5] The linearity of the relationship between internal friction and the weight fraction of solute atoms has been determined by L. J. Dijkstra, Philips Res. Report 2, 357 (1947).

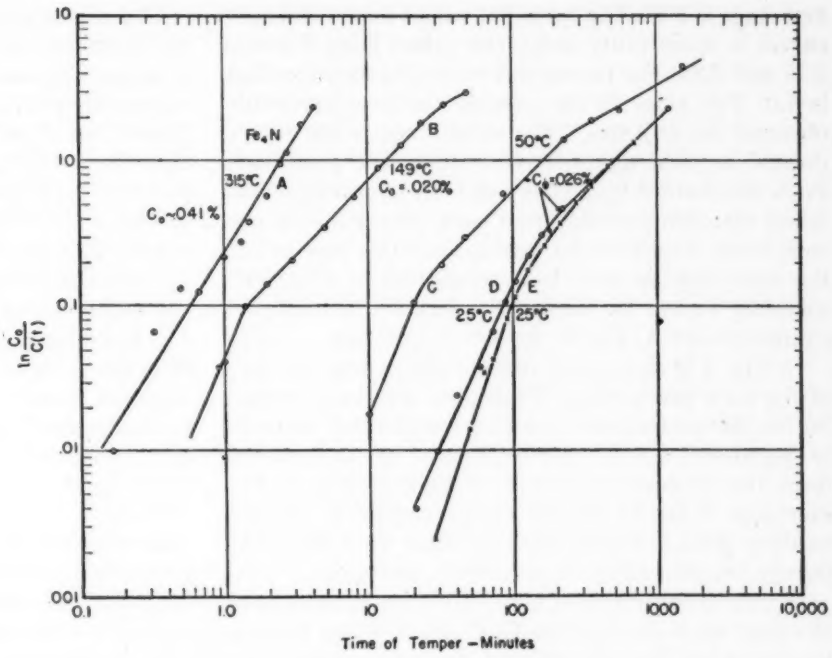


FIG. 3. Precipitation of nitrogen. Curves B, C, D, and E are data on precipitation of N Phase 1; curve A for Fe_4N.

diameter, no heat treatments were made. Since it was desirable from the experimental point of view to use total tempering times of the order of hours or days, the range of concentration and temperature was somewhat limited. For Fe_3C measurements were made over the range 35°C to 315°C with an initial carbon concentration of about 0.016 wt. percent; for N Phase 1, 25°C to 150°C with an initial nitrogen concentration of 0.02 to 0.026 wt. percent; and for Fe_4N, 315°C with an initial concentration of 0.042 wt. percent. These initial concentrations were determined by measuring the height of the internal friction peak immediately after the specimen was quenched.

The attempt to fit the experimental data with a transformation equation was quite successful, especially for the precipitate Fe_3C. The equation used has the same form as one developed by Johnson and Mehl[6] to characterize crystal growth. The development of this equation may be seen from the following considerations: Suppose that particles of a second phase grow in a matrix without mutual interference. The rate of the transformation may be written

$$(dW(t)/dt)=f(t),$$

where $W(t)$ is the fraction of transformed material and $f(t)$ is some function of time. But because of mutual interference this rate will be lessened; the decrease will be proportional to the fraction already transformed. Hence, with the assumption of random spatial distribution of the particles the rate of transformation may be written

$$(dW(t)/dt)=[1-W(t)]f(t).$$

Solving this equation gives

$$W(t)=1-\exp\left[-\int_0^t f(t)dt\right].$$

The explicit form of this equation used was

$$\Delta C(t)/C_0=1-\exp[-(t/\tau)^n], \qquad (1)$$

where $\Delta C(t)$ is the amount of transformed material, C_0 is the initial concentration of solute atoms, τ is a time constant depending on C_0, the temperature, T, and some other undetermined factors. In this equation, n is some number independent of T and C_0. Equation (1) may be expressed in the form

$$\log \ln[C_0/C(t)]=n \log t-n \log \tau, \qquad (2)$$

where $C(t)$ is the concentration in solution at time t; $C_0=C(t)+\Delta C(t)$.

The data obtained for the precipitation of Fe_3C at various temperatures are given in Fig. 2, in which the data are plotted according to Eq. 2. In this plot, the slope of the lines determines n, and the intercept on the line $\ln[C_0/C(t)]=1$ determines τ. The value of n obtained by averaging the slopes of these seven curves was $n=1.45$. τ varies from about 12,000 minutes at 43°C to ½ minute at 312°C. For the different curves, n varies between 1.2 and 1.7.

Similar data obtained for the precipitation of N Phase 1 and Fe_4N are shown in Fig. 3. The curves for N Phase 1 are seen to consist of three distinct regions. The first stage extends to amounts of precipitate up to about 20 percent of the total precipitate which will form; the second, to about 80–90 percent; and the third, the remainder. The average value of n for the

[6] W. A. Johnson and R. F. Mehl, Trans. Am. Inst. Min. Met. Eng. 135, 416 (1939).

first stage is 2.45. The spread of values for the different curves is again fairly large, the values lying between 2.25 and 2.85. For the second stage, the average slope is 1.0. This stage in the process has been previously observed by Dijkstra.[6] The third stage is too poorly defined to yield quantitative results. The precipitate Fe_4N was formed by a choice of both temperature and initial nitrogen concentration such that Phase 1 was completely soluble at all times. That this was indeed the case may be seen by examination of Dijkstra's solubility curves for these two phases.[2] For this precipitate (curve A, Fig. 3), $n=1.6$; $\tau=2.6$ min.

In Fig. 4 is shown one precipitation curve for each of the three precipitates. These plots are made according to the coordinates "percent transformed material *vs.* log time." τ and n for each curve are values read from the appropriate curves in Figs. 2 and 3. The exactness of the fit for the carbon precipitate is particularly good, even for very long times when the interference between adjacent particles is very great.

In view of the nature of these experiments, the spread of values for n obtained for Fe_3C and N Phase 1 is not too surprising. The difficulties of measurement have two sources in general. Firstly, during the early stages of the precipitation process small errors in the determination of C_0 and $C(t)$ are reflected as errors of rather large percentage in $\ln[C_0/C(t)]$. Secondly, there is the difficulty of knowing the exact tempering time, especially at the higher temperatures. The time of heating and quenching may not be small compared to the total measured tempering time, even though care was taken to reduce this possibility.

One of the objectives in the study of precipitates is to determine their physical nature. In this work an attempt was made to determine the shape of the precipitate simply from analysis of the precipitation curves. Zener[7] has shown the following to be true: Assume that the rate of growth of a particle of a second phase in a matrix of infinite extent is controlled by a diffusion process. Then for precipitates of different physical shapes, it is possible to predict the time dependence of the transformed material. For a sphere it is found that the radius increases as $t^{\frac{1}{2}}$, hence $V_{\text{sphere}}\sim(t-t')^{\frac{1}{2}}$, where t' is the time at which the nucleus was formed. For a disk the thickness also increases as $t^{\frac{1}{2}}$, but the radius increases linearly with time since the growth conditions at the advancing edge of the disk remain constant. Hence for this case $V_{\text{disk}}\sim(t-t')^{5/2}$. For a cylinder, the radius increases as $t^{\frac{1}{2}}$ and the length as t, so $V_{\text{cylinder}}\sim(t-t')^{3/2}$.

To see how the experimental work may be interpreted in the light of these observations, we see that for times such that $t\ll\tau$, Eq. 1 may be written $\Delta C(t)/C_0=(t/\tau)^n$; the same form as in Zener's observations. Now the average values of n obtained by experiment for precipitation of Fe_3C is about $\frac{3}{2}$, and for the initial stage in the formation of N Phase 1, about 5/2. Comparing these experimentally determined values with the values of n obtained by the analysis mentioned in the preceding paragraph, we see that the data may be interpreted as showing that: (1) Fe_3C precipitates in the form of spheres and (2) N Phase 1 precipitates initially in the form of disks. This interpretation sets $t'=0$; i.e., all nuclei which ultimately grow are of the

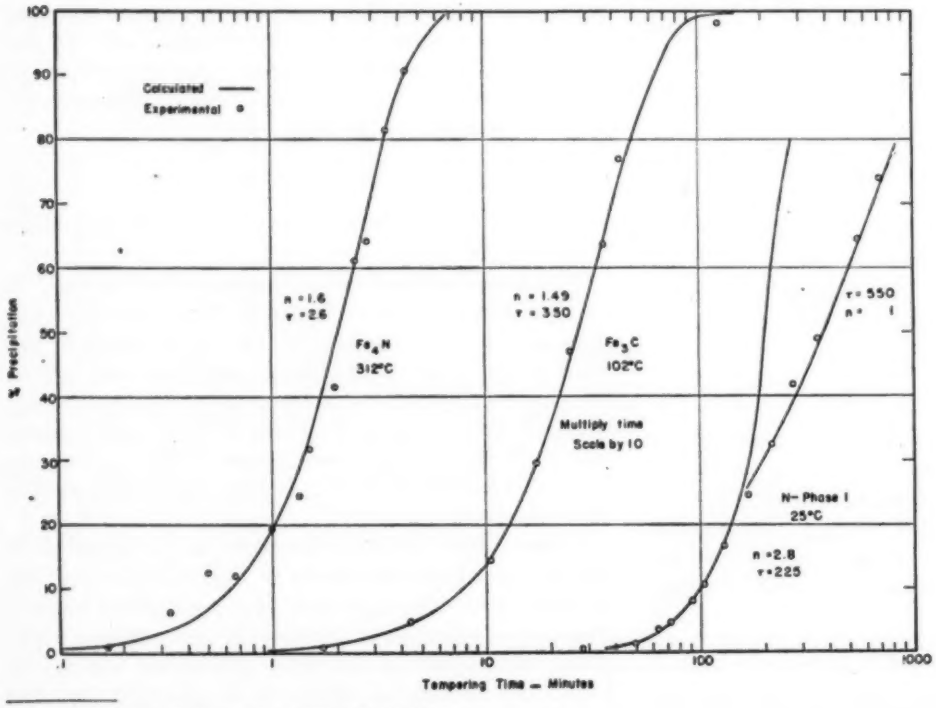


FIG. 4. Typical precipitation curves plotted as amount of precipitation *vs.* time. The smooth curves are calculated from the equation $\Delta C(t)/C_0=1-\exp(-t/\tau)^n$ with n and τ determined from the appropriate curves of Figs. 2 and 3.

[7] See C. Zener, J. App. Phys. **20**, 950 (1949).

athermal variety and are present immediately after quenching.[8] There is consequently no incubation period involved in the process.

At first sight it may appear that this interpretation given to the experimental data is somewhat arbitrary; i.e., would it not be possible to satisfy the experimental data with some other physical shape and thereby allow a process of continuous nucleation? The answer to this question depends on the following arguments: If nucleation occurs throughout the precipitation process according to some power of t, say t^s, and the particle growth as t^r, then the net rate of precipitation is t^{r+s}. We have seen that possible values of r are 3/2, 4/2 and 5/2 for three common geometrical shapes. Further, since any face of any geometrical figure must grow at least as rapidly as $t^{\frac{1}{2}}$, it follows that the minimum value which r can assume for unhindered growth is 3/2. With these considerations examine the interpretation given the data. It is clear that for Fe_3C there is no alternative; it must precipitate as spheres. For N Phase 1 the evidence is not so strong that disks are formed, for the growth of spheres or cylinders combined with an appropriate nucleation rate could likewise give $n=5/2$. It will be shown in the last section, however, that if there is continuous nucleation, a further coincidence is required, namely that the nucleation rate should depend on temperature through a factor $\exp(-H_N/RT)$ where H_N is the heat of activation associated with the interstitial diffusion of N in α-iron. That these two coincidences should simultaneously be satisfied seems unlikely; hence the interpretation of the author is that N Phase 1 precipitates as disks.

It is, of course, apparent that even with the choice of spheres and disks for Fe_3C and N Phase 1, respectively, that some continuous nucleation may occur. The data are not good enough to exclude this possibility, especially during the early stages of the process. The most that can be claimed is that a large part of the precipitate forms around nuclei which are present when the specimen is quenched or which appear shortly afterward.

No interpretation is given to the data for Fe_4N. The inaccuracy involved in the extremely short tempering times makes an interpretation too uncertain to be meaningful. The precipitate first becomes visible under a microscope when the precipitation is 30 percent complete. At this stage it has the appearance of short stubby plates. These plates grow and coalesce with time. Observations made earlier on Fe_4N at much higher concentrations and temperatures showed also that Fe_4N precipitates in plates.[9]

It is interesting to examine the type of interference which is present in the cases considered here. For the most part the interference is of a different nature than would be found, for example, in recrystallization, where the growing crystals directly impinge upon each other. Rather, as the matrix around the particles becomes depleted of solute atoms, the different adjoining particles begin to draw atoms from the same region. This lowers the rate of flow of solute atoms to all particles concerned and retards their growth. In the instance of Fe_3C the interference correction in Eq. 1 just accounts for this decrease, but not so for the instance of N Phase 1. Here the disks of precipitate interfere at a very early stage in the process. The interference is quite abrupt and produces a distinct lowering of the precipitation rate. These facts are interpreted as showing that the individual neighboring disks do not lie parallel to one another, although it is believed that they do form on a definite set of crystallographic planes. This non-parallelism results in their "growing into" each other while much of the nitrogen still remains in solid solution. When this occurs, the rates of growth in both thickness and in radius change radically and the growth of the particles begins to obey a different law.

Metallographic examination of these precipitates showed the following: Below 50°C no precipitates are ever seen for these concentrations of solute atoms for tempering times of several weeks. For N Phase 1, the specimen used in making curve B, Fig. 3 was examined microscopically at 12 stages during the precipitation. No precipitate was visible for tempering times below five minutes. For tempering times longer than this, small plates became visible which coalesced into larger plates apparently on 100 planes. It is thought likely that the second bend in curve B is caused in part by this coalescence and by a small amount of coalescence in some grain boundaries. This would tend to hold the amount of nitrogen in solid solution at a higher level than otherwise. For Fe_3C at 312°C (curve A, Fig. 1) some small dots of precipitate are visible at all times exceeding 20 seconds. These particles grow in size and become more numerous in a polishing plane due to growth. No coalescence is observed to tempering times of 18 minutes. Stubby plates of Fe_4N (curve A, Fig. 3) became visible after one minute. These plates lie on planes of higher index. As has been mentioned previously, they do not seem to be entirely random; occurring in rows which seemed to lie along oxide stringers.

NUCLEATION

Figures 2 and 3 show that the rate of precipitation is strongly temperature dependent. We wish now to express this fact in Eq. (1). This obviously must be done by making τ a function of temperature. For temperatures far below the equilibrium temperature, this dependence may have two sources. These are the rate of diffusion of the solute atoms and the distance over which they must travel to reach the precipitate. Then

$$\tau = L^2/D,$$

where D is the diffusion coefficient and L is propor-

[8] Fisher, Holloman, and Turnbull, J. App. Phys. **19**, 775 (1948).
[9] Mehl, Barrett, and Jerabek, A.I.M.E. Iron and Steel Division, Tech. Pap. 539 (1934).

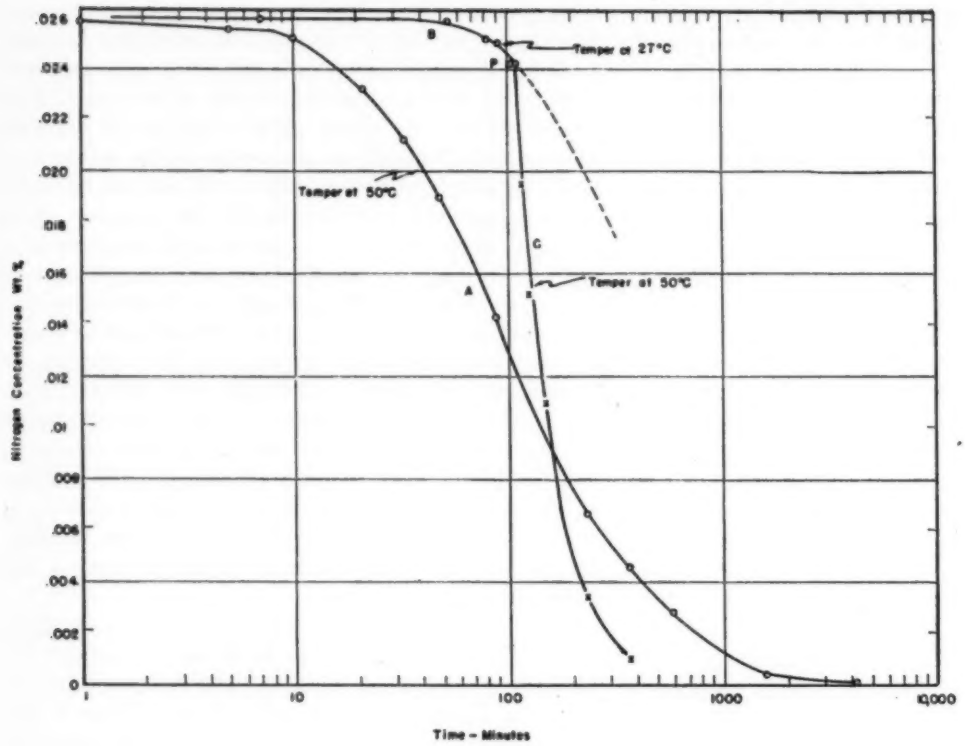


FIG. 5. Influence of number of nuclei on precipitation rate.

tional to the mean distance of diffusion and is hence related to the mean distance between the particles of precipitate. Since $D=D_0e^{-H/RT}$ we may write

$$\tau=(L^2/D_0)e^{+H/RT},$$

where H is the heat of activation associated with the interstitial diffusion of the solute atoms and R is the gas constant. The dependence of L on temperature is not known; however, let us suppose that one can devise an experiment which will keep L constant over a moderate temperature range. Under these conditions Eq. (1) becomes

$$\Delta C(t)/C_0=1-\exp[-D_0te^{-H/RT}/L^2]^n.$$

To determine H one observes the time required for the same degree of precipitation at different temperatures. Then

$$[d\Delta C(t)/C_0/d(1/T)]=0,$$

and it follows that

$$H=2.3R[d\log t/d(1/T)]. \quad (3)$$

An early calculation based upon some of the data shown in Figs. 2 and 3 demonstrated that L is certainly dependent on temperature. This computation was carried out by first finding at different temperatures the time required for the precipitation to become one-half complete. Then $\log t$ was plotted *vs.* $1/T$ and the slope of this line determined. From this slope H was determined using Eq. (3). The values obtained were $H_N=14{,}000$ cal./mole and $H_C=13{,}000$ cal./mole. These values are considerably lower than the published values for these constants of 20,000[3] and 18,400,[1,10] respectively. This deviation shows that L must be highly temperature dependent; i.e., that the number of nuclei effective per unit volume must depend markedly on temperature. To investigate this fact the following experiment was carried out: Precipitation of N was carried out at 50°C. This is shown by curve A, Fig. 5. The precipitate was then redissolved by heating and quenching. Precipitation at 50°C was again carried out, but now was preceded by a precipitation period at 27°C, in which 6 percent of the nitrogen was precipitated. This treatment formed particles which apparently were stable at 50°C and which served as precipitation centers at 50°C. It is apparent from comparison of curves A and C that more nuclei were present at 27°C than were present at 50°C.

In passing it is fitting to make two remarks about this experiment. (1) It is seen from Figs. 2 and 3 that precipitation takes place more rapidly as the temperature increases. This is caused solely by the enormous increase of diffusion rate with temperature which more than overwhelms the greater distance the atoms must diffuse to reach a precipitation center. (2) This "seeding" of a second phase can be very useful. In making solubility measurements, the author has used this device to shorten considerably the time necessary to attain equilibrium between two phases.

At this point it seemed likely that one might also eliminate the dependence of L on temperature by this technique. Accordingly, precipitation experiments on

[10] The author has recently carried out a redetermination of H_C by the method discussed in reference 3. The value obtained for H_C was 19,800±400 cal./mole.

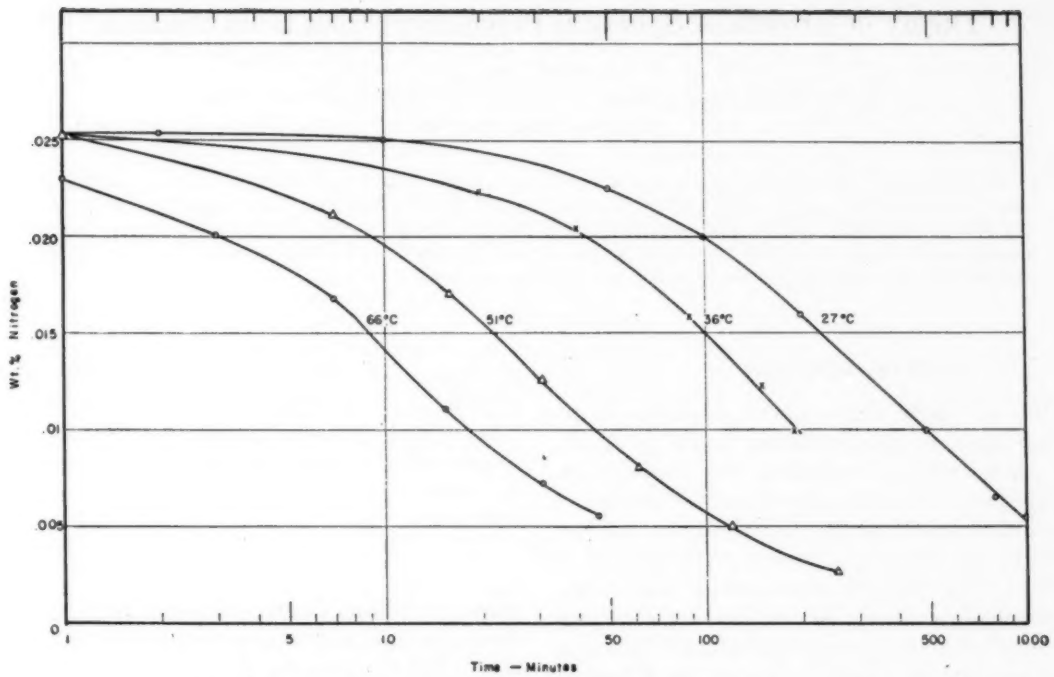


FIG. 6. Precipitation at various temperatures following nucleation at 27°C.

N were made at a number of temperatures following nucleation at 27°C. These data are shown in Fig. 6. For each curve the time $t=0$ and the value of C_0 were chosen as those values of t and $C(t)$ at the end of the nucleation period at 27°C. This is illustrated by point P in Fig. 5. The data in Fig. 6 were then used to compute H_N using Eq. 3 and the method outlined in a preceding paragraph. The average value for H_N obtained by this procedure was 18,000 cal./mole±10 percent. A similar calculation carried out in the range 100°C to 150°C for carbon gave $H_C=17{,}500\pm10$ percent. These values agree within experimental error with the accepted values of H_N and H_C given above.

The results of this experiment can be shown to impose a condition on the temperature dependence of any continuous nucleation which may occur. This is shown by the following arguments: The amount of precipitate at any time may be written

$$W\sim\int_0^t f[(t-t')\exp(-H_N/RT)]\times\frac{\partial}{\partial t'}[t'\exp(-H'/RT)]^a dt', \quad (4)$$

where the function f is the growth law of a precipitate of a given shape, $\partial/\partial t'[t'\exp(-H'/RT)]^a$ is the assumed nucleation rate and t' is the time at which a nucleus is formed. Now, from the preceding paragraph it is seen that the measured dependence of W on time and temperature is given by

$$W(t,T)=W[t\exp(-H_N/RT)].$$

It can readily be shown in Eq. (4) that this imposes the condition that $H'=H_N$. It is this requirement which is cited previously in the discussion of the nature of the N Phase 1 precipitate.

In the preceding section, some observations of Zener on growth rates were mentioned. His later considerations of the problem have given the proportionality constants enabling him to give an explicit expression for the volume of precipitate.[7] For the sphere this is

$$V_{\text{sphere}}=(4\pi/3)\alpha^3(Dt)^{\frac{3}{2}}, \quad (5)$$

where D is the diffusion coefficient, and α is a growth coefficient depending on certain concentration constants. With the aid of this expression, the data shown in Fig. 2 can be analyzed to give the number of nuclei effective at each temperature. From experiment, the amount of precipitate which formed in a certain period of time was determined. The size of each sphere of precipitate at that time was computed using Eq. (5). The quotient of these two quantities then gave the number of nuclei which formed and grew at that temperature. These calculations, of course, were carried out at an early stage of the precipitation process; before interference between adjacent particles became important.

Such computations for Fe_3C gave the following results for initial carbon concentrations of about 0.016 wt. percent. At 37°C, the number of nuclei per cc was 2×10^{15}; at 102°C, 10^{14}; and at 312°C, 3×10^{12}. On the average then, the distance between adjoining nuclei at 37°C, 102°C and 312°C was about 800, 2000, and 7000 angstroms, respectively.

The author would like to express his great appreciation to Dr. Zener for many valuable discussions during the course of the investigation.

Theory of Growth of Spherical Precipitates from Solid Solution*

CLARENCE ZENER
Institute for the Study of Metals, The University of Chicago, Chicago, Illinois
(Received March 14, 1949)

The radius of a spherical precipitate particle growing in a solid solution of initially uniform composition may be shown to be equal to $\alpha(Dt)^{\frac{1}{2}}$, where D is the atomic diffusion coefficient, t the time of growth, and α, the growth coefficient, is a dimensionless function of the pertinent compositions. In this paper the precise dependence is found of this function upon the pertinent concentrations. A similar computation is made for the growth coefficient corresponding to the one-dimensional growth of a plate.

A. INTRODUCTION

DURING the early stages of the experiments reported upon in the preceding paper by Wert[1] upon the precipitation of spherical iron carbide particles from α-iron, it became clear that the amount of information which could be deduced from the observations would be increased by a theory of the growth of such particles. The present investigation was undertaken to supply this theory. As finally developed, this theory is applicable to the growth of any spherical precipitate the rate of which is limited by diffusion. It has been found possible to generalize to n dimensions an analysis previously given by Dube[2] for the one-dimensional problem of the sidewise growth of a precipitate plate. The present study of this more general problem has made it possible to present the theory of the one-dimensional problem in a more complete form than has been given heretofore.

The results of this analysis are reviewed briefly below. Let S refer to the growth coordinate. Thus in the one-dimensional case S is the half-width of the plate, in the case of the two-dimensional radial growth of a cylinder S is the radius of the cylinder, in the three-dimensional case of a spherical particle S is the radius of the particle. From dimensional arguments alone it is shown that S varies with time according to the equation

$$S=\alpha_\lambda(Dt)^{\frac{1}{2}}, \tag{1}$$

where D is the atomic diffusion coefficient, and the dimensionless coefficient α_λ, hereafter called the growth coefficient, is a function of the pertinent concentrations. The suffix λ refers to the number of dimensions. The pertinent concentrations n_0, n_1, n_∞ are illustrated in Figs. 1 and 2. The first figure illustrates the case of a positive precipitation, e.g., iron carbide in α-iron, where the solute atoms must diffuse into the moving interface. The second figure refers to the case of a negative precipitation, e.g., the growth of a ferrite particle in austenite, where the solute atoms must diffuse away from the moving interface. In both cases n_0 is the concentration, referred to unit volume, of the solute atoms in the precipitate, n_1 is that concentration in the matrix which is in equilibrium with the precipitate, and n_∞ is the concentration in the matrix far away from the precipitate. N_∞ lies between n_0 and n_1. The growth coefficient α_λ depends only upon the relative value of n_∞ with respect to n_0 and n_1. Thus, in the numerical evaluation α_λ is found as a function of the parameter

$$\xi=(n_\infty-n_1)/(n_0-n_1). \tag{2}$$

It is found, however, that α_λ is not a simple function of this parameter, varying from 0 to ∞ as this parameter varies from 0 to 1. The most natural parameters are found from semiquantitative considerations in Section C. In terms of these natural parameters it is found that in the one-dimensional case of $\lambda=1$,

$$\alpha_1=K_1\frac{(n_\infty-n_1)}{(n_0-n_\infty)^{\frac{1}{2}}(n_0-n_1)^{\frac{1}{2}}}, \tag{3}$$

where K_1 varies slowly from 1.13 to 1.41 as n_∞ varies from n_1 to n_0. The precise dependence of α_1 upon concentration is given in Fig. 3. In the three-dimensional case of $\lambda=3$ it is found that

$$\alpha_3=K_3(n_\infty-n_1)^{\frac{1}{2}}/(n_0-n_\infty)^{\frac{1}{2}}, \tag{4}$$

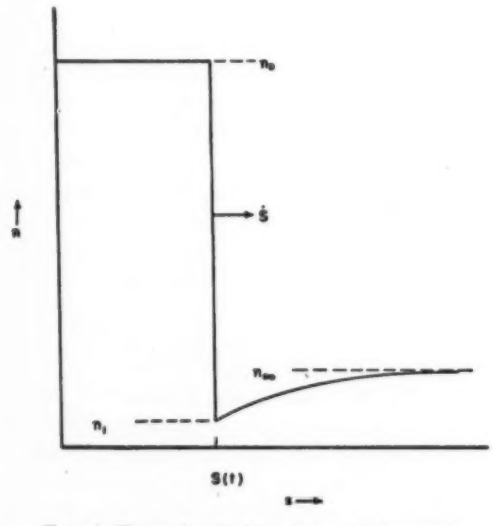


FIG. 1. Example of advancing interface with positive precipitation.

* This research has been partially supported by the ONR, U. S. N. (Contract No. N-6ori-20-IV, Contract NR o15 o18).
[1] C. Wert, J. App. Phys. **20**, 943 (1949).
[2] Arthur Dube, doctorate thesis, 1948, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

where K_3 varies slowly from 1.4 to 2.4 as n_∞ varies from n_1 to n_0. The precise dependence of α_3 upon concentration is given in Fig. 4.

B. ANALYSIS

We shall consider that the atomic diffusion coefficient D is independent of concentration. The concentration of solute atoms, n, will then satisfy the standard diffusion equation

$$\partial n/\partial t = Ds^{-\lambda+1}(\partial/\partial s)s^{\lambda-1}(\partial n/\partial s). \tag{5}$$

In the case of a plate precipitate ($\lambda=1$) s is the linear coordinate, while in the case of a spherical precipitate ($\lambda=3$), s is the radial coordinate.

The boundary conditions appropriate to our problem are of three types. We require firstly that $n(t,s)$ satisfy

$$n(0,s)=n_\infty, \tag{6}$$

corresponding to no precipitate being initially present. The second boundary condition refers to the concentration at the interface. We shall neglect the effect of surface tension upon the equilibrium concentration. Our final results will, therefore, be valid only after the precipitate particles have attained a size large compared to the critical size of a stable nucleus. We shall make the further assumption that the rate of growth

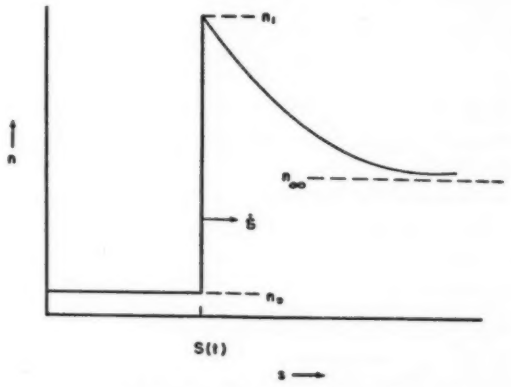


FIG. 2. Example of advancing interface with negative precipitation.

is limited solely by diffusion, i.e., that the interface moves just as rapidly as diffusion permits. According to these approximations the boundary condition at the interface is

$$n\{t, S(t)\}=n_1, \tag{7}$$

where $S(t)$ is the value of s at the interface. The final boundary condition relates the rate of advance of the interface, $\dot{S}$, to the gradient of the concentration at the interface. Upon equating the rate at which solute atoms are left behind the interface in the precipitate to the rate at which the solute atoms encounter the interface in the matrix, we obtain

$$n_0\dot{S}=n_1\dot{S}+D(\partial n/\partial s)_{s=S}, \tag{8}$$

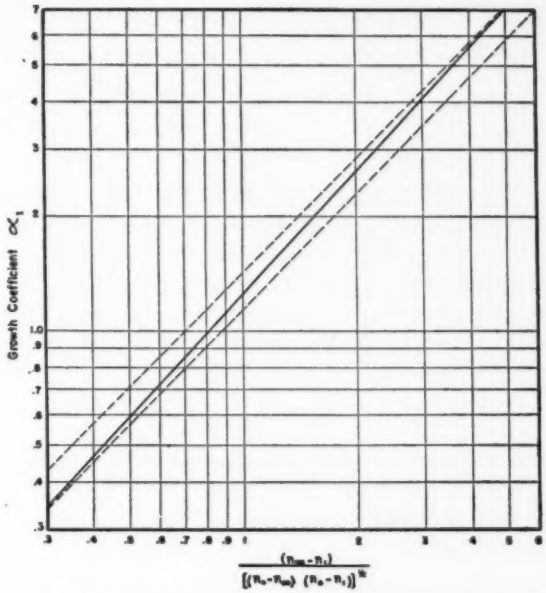


FIG. 3. Growth coefficient for one-dimensional growth.

or

$$(n_0-n_1)\dot{S}=D(\partial n/\partial s)_{s=S}. \tag{9}$$

Since both the governing differential equation and all the boundary conditions are homogeneous in the concentrations, the rate of advance $\dot{S}$ can depend upon the concentrations n_0, n_1 and n_∞ only in a dimensionless combination. The only remaining quantities upon which $S(t)$ can depend are D and t. Dimensional arguments, therefore, lead to the general form of Eq. (1) for $S(t)$.

That solution of Eq. (5) which satisfies the boundary conditions (6)–(7) is

$$n=n_\infty+(n_1-n_\infty)\phi_\lambda[s/(Dt)^{\frac{1}{2}}]/\phi_\lambda(\alpha_\lambda), \tag{10}$$

where

$$\phi_\lambda(x)=\int_x^\infty e^{-\frac{1}{4}z^2}z^{1-\lambda}dz, \tag{11}$$

and α_λ is defined by (1). The function $\phi_1(x)$ is a standard solution of the one-dimensional diffusion equation. As far as the author is aware, the generalized solution $\phi_\lambda(x)$ to the multi-dimensional diffusion equation has not been heretofore presented. That it is a solution may, of course, be verified by direct substitution. The third boundary condition (9) gives us the following equation for the growth coefficient

$$\xi^{-1}\cdot(\alpha_\lambda)^\lambda=2\exp(-\alpha_\lambda^2/4)\cdot\phi_\lambda^{-1}(\alpha_\lambda), \tag{12}$$

where ξ is defined in Eq. (2). This equation has previously been given by Dube for the case $\lambda=1$.

Equation (12) uniquely determines the growth coefficient in terms of the parameter ξ. The relation between α_λ and ξ in the two extreme cases of very small and of very large values of α_λ may be obtained by means

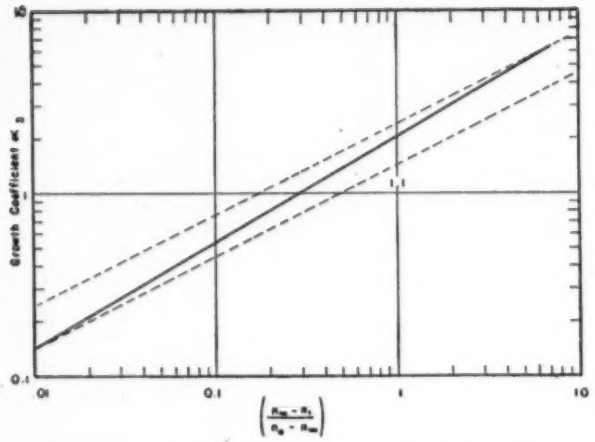


FIG. 4. Growth coefficient for three-dimensional growth.

of the two asymptotic relations[3]

$$\phi_1(\alpha)\simeq\pi^{\frac{1}{2}}, \quad \alpha\ll 1,$$

$$\phi_1(\alpha)\simeq(2/\alpha)\exp(-\alpha^2/4)(1-2\alpha^{-2}+12\alpha^{-4}+\cdots), \quad \alpha\gg 1,$$

and of

$$\phi_3(\alpha)=\alpha^{-1}\exp(-\alpha^2/4)-(\tfrac{1}{2})\phi_1(\alpha),$$

the latter relation being obtained by integration by parts. These asymptotic solutions are

$$\alpha_1=\begin{cases}2\pi^{-\frac{1}{2}}\xi, & \alpha\ll 1\\ 2^{\frac{1}{2}}(1-\xi)^{-\frac{1}{2}}, & \alpha\gg 1,\end{cases} \tag{13}$$

and

$$\alpha_3=\begin{cases}2^{\frac{1}{2}}\xi^{\frac{1}{2}}, & \alpha\ll 1\\ 6^{\frac{1}{2}}(1-\xi)^{-\frac{1}{2}}, & \alpha\gg 1.\end{cases} \tag{14}$$

The relation between α_λ and ξ in the intermediate ranges may readily be obtained by graphical methods. One first takes the λth root of Eq. (12), and then makes a graph of the right-hand member as a function of α_λ. The abscissa of the intersection of the straight line $\xi^{-1/\lambda}\alpha_\lambda$ with this graph then gives the value of α_λ associated with ξ. In presenting the results of this graphical solution, use will be made of the concentration parameters which are found in the following section to be more appropriate to our problem than the parameter ξ. In terms of these new parameters the asymptotic solutions (13) and (14) become Eqs. (3) and (4). The results are presented in Figs. 3 and 4. The asymptotic solutions (13) and (14) are represented as dashed lines. The actual solution is seen to pass smoothly from one asymptotic solution to the other in the intermediate region of α_λ.

C. APPROXIMATE SOLUTION

In Section B a rigorous solution was obtained for the velocity of advance $\dot{S}$ in terms of the concentration parameter ξ defined by (2). In the present section an approximate solution is presented for the sole purpose of obtaining a concentration parameter more suitable than ξ for expressing the exact results of the previous section.

The method of approach will be to estimate the value of $\partial n/\partial s$ at the interface, and then to substitute this estimate into the boundary condition (9) and then solve for the function $S(t)$. In the one-dimensional case we shall estimate the value of $\partial n/\partial s$ at the boundary by assuming it to have that value corresponding to the approximate solution represented in Fig. 5. In this solution n is taken to change at a constant rate from the value n_1 to n_∞, the rate being determined by the condition that the total number of solute atoms remains constant. This condition requires the equality of the two shaded regions. We thereby obtain the following estimate

$$\partial n/\partial s|_{s=S}\simeq\Delta n/\Delta s,$$

where

$$\Delta s=2(n_0-n_\infty)S/\Delta n,$$

and

$$\Delta n=n_\infty-n_1.$$

Upon substituting this estimate back into Eq. (9), and solving for S, we obtain

$$S\simeq\alpha_1^*(Dt)^{\frac{1}{2}},$$

with

$$\alpha_1^*=\frac{(n_\infty-n_1)}{(n_0-n_1)^{\frac{1}{2}}(n_0-n_\infty)^{\frac{1}{2}}}.$$

The quantity α_1^* will therefore be a natural parameter in terms of which to express the exact solution of the one-dimension problem.

A corresponding estimate of the growth coefficient for a spherical particle can be obtained only in two extreme cases. In the first extreme case, corresponding to the inequality

$$(n_0-n_\infty)/(n_0-n_1)\ll 1, \tag{15}$$

the solute atoms are depleted from only a comparatively thin shell surrounding the spherical precipitate. Our estimate of the growth coefficient then proceeds

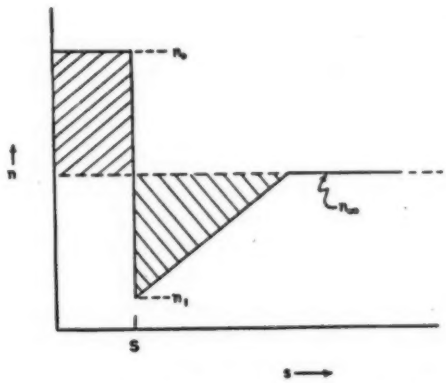


FIG. 5. Assumed approximate solution in one-dimensional case.

[3] E. Jahnke and F. Emde, *Funktionentafeln* (Dover Publications, New York, 1943), p. 24.

precisely as in the one-dimensional case, leading to

$$\alpha_3^* = 3^{\frac{1}{2}}\alpha_1^*. \tag{16}$$

In the second extreme case, corresponding to the condition

$$(n_\infty - n_1)/(n_0 - n_1) \ll 1, \tag{17}$$

depletion of the solute atoms extends to a distance large compared to the radius of the precipitate particle. In this case we shall estimate the value of $\partial n/\partial s$ at the interface by assuming it to have that value corresponding to the steady-state solution in which the boundary is stationary, namely the solution

$$n = n_\infty - (n_\infty - n_1)(S/s).$$

We thereby obtain

$$\alpha_3^* = 2^{\frac{1}{2}}\{(n_\infty - n_1)/(n_0 - n_1)\}^{\frac{1}{2}}. \tag{18}$$

Now under the conditions of the inequality (15), α_3^* as given by Eq. (16) is identical to α_3^* as given by Eq. (18), aside from the factor of $(3/2)^{\frac{1}{2}}$. The quantity α_3^* as given by Eq. (18) will, therefore, be a natural parameter in terms of which to express the exact solution of the expanding spherical precipitate particle.

Infra-Red Spectra and Structure of Hevea and Gutta Elastomers

R. A. Saunders and D. C. Smith
Naval Research Laboratory, Washington, D. C.
(Received March 11, 1949)

The infra-red absorption spectra of the purified rubber hydrocarbon fractions of several naturally occurring elastomers (hevea, guayule, kok-saghyz, gutta-percha, balata, and chicle) have been determined at room temperature over the range 2–15μ and it has been shown that the elastomers can be classified on the basis of the spectral data as belonging to one or the other of two different types, namely the hevea type or the gutta type. Both types of elastomers have also been studied over the temperature range −25 to +80°C and reference spectra for both the amorphous and the solid (crystalline) phase(s) occurring in this range have been determined for each type. At room temperature the spectral differences between the hevea (amorphous state) and gutta (α- or β-crystalline modifications) types are much more pronounced than is the case when the spectra of both types in the same (amorphous) states are compared.

The spectra of two pair of *cis-trans*-isomers containing the structural unit of polyisoprene have been determined and correlated with the elastomer spectra. The results indicate but do not prove that hevea- and gutta-type elastomers have the molecular structure of *cis*- and *trans*-polyisoprene, respectively.

The infra-red absorption spectra of oriented crystalline films of hevea rubber and of gutta-percha in both the α- and β-forms have also been obtained over the range 2–15μ using polarized radiation. The observed dichroic effects lead directly to the conclusion that hevea- and gutta-type elastomers are *cis*- and *trans*-polyisoprene, respectively. Although it has not been possible to deduce the exact spacial configuration of the molecules in the crystallites, it has been indicated that the planar isoprene models are unsatisfactory and that the results are in close agreement with the twisted models proposed on the basis of x-ray diffraction.

Further use has been made of the polarization data in proposing nearly complete but tentative interpretations of the spectra of both hevea- and gutta-type elastomers. The possibility of making use of the vibrational assignments for the interpretation of spectral changes produced by oxidation or other forms of degradation is pointed out.

INTRODUCTION

In a previous publication[1] Dinsmore and Smith have reported spectral data for natural (hevea) rubber and for several synthetic elastomers and have discussed in some detail the applicability of spectral methods to the determination of molecular structure and the analysis of elastomer mixtures. In an extension of this work the spectrum of hevea has been re-examined and five other naturally occurring elastomers have been studied in both amorphous and crystalline form and with polarized radiation. The newer results together with certain observations of basic importance to the application of spectral methods to rubber research are presented here.

DESCRIPTION OF SAMPLES

Samples of hevea (smoked sheet), guayule, kok-saghyz, gutta-percha, balata, and chicle as originally obtained[2] were crude and uncompounded plantation products. The crude guayule was in the form of "worms" prepared by the mechanical separation of the elastomer from plants grown in California. It had been stored under water from the time of preparation and was light brownish green in appearance. All other samples had previously been stored in the dark but their exact age was unknown. An additional sample of guayule, a compounded specimen prepared in the rubber laboratories at Salinas, California, and a refined product of gutta-percha which was clean and white as compared to the dirty brown appearance of the crude sheets, were also examined.

In general the samples were prepared for spectral examination by the procedures recommended previously.[1] Films of unextracted elastomer were prepared by dissolving the original sample in benzene at room temperature and in the dark, filtering and/or centri-

[1] H. L. Dinsmore and D. C. Smith, Anal. Chem. **20**, 11 (1948).

[2] From the Plant Industry Station, Beltsville, Maryland, and from the National Bureau of Standards, Washington, D.C.

fuging to remove dirt, sediment, and insoluble or gel rubber, and evaporating in successive layers a definite volume of 1.0 percent solution within a prescribed area on an NaCl or KBr plate. Purified samples were prepared by extraction with acetone followed by at least one precipitation from a clear benzene solution using acetone as the precipitant.

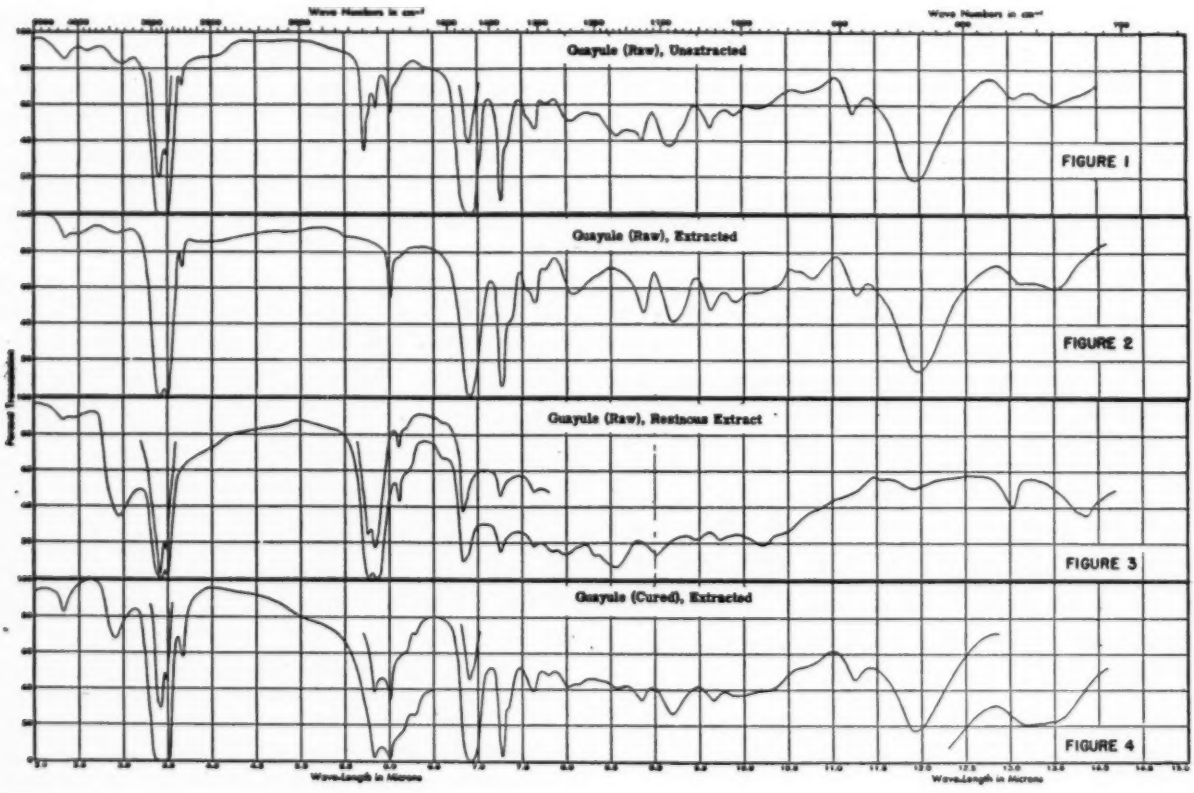
All spectra in this work were obtained with the large prism spectrophotometer employed previously.[3] All films were examined over the 2–15-μ-range using the NaCl prism. For measuring spectra above 25°C a high temperature cell similar to that described by Richards and Thompson[4] was used. For measurements below room temperature (−25°C) the film plate was placed in a metal cell block, cooled by a petroleum base heat-transfer fluid circulated through the cell and through cooling coils immersed in a dry-ice acetone bath. The cell block was mounted in an evacuated chamber fitted with KBr windows and the temperature was measured by a thermocouple attached to the specimen plate.

IDENTIFICATION OF HEVEA- AND GUTTA-TYPE ELASTOMERS

The spectra of the unextracted elastomers, the purified elastomers, and the resinous extracts are shown in Figs. 1 to 16 for all specimens except hevea. These spectra were obtained at room temperature. They show that in each case the sample has been separated into a fraction consisting of the pure rubber hydrocarbon and a non-rubber fraction, which, in some cases, contains a small amount of rubber due to the action of extraction solvent. Some of the acetone extracts precipitated white solid spheroids upon cooling and this material was shown by spectral examination to consist of rubber hydrocarbon, probably of low molecular weight. In each case, however, the resin and rubber hydrocarbon are readily distinguished by infra-red absorption.

The spectrum of hevea at room temperature is shown in Fig. 17 (solid curve). It differs appreciably from the spectra of crude guayule (Fig. 1) and crude kok-saghyz (Fig. 5), but most of the differences are due to resinous material in the latter products. The spectra of the purified elastomers (Figs. 2, 6, and 17) are very nearly identical (when allowance is made for differences in film thickness) and serve to confirm the belief that the polymeric units are the same for these three elastomers, which are therefore called hevea-type elastomers.

Crude samples of gutta-percha, balata, and chicle also contain considerable quantities of resin as shown clearly by comparison of the spectra of the crude and purified samples. However, when the resins are removed and the purified elastomers are examined in the same physical state, the spectra obtained (Figs. 9, 12, and 15) are nearly identical and establish the equiva-



[3] Nielsen, Crawford, and Smith, J. Opt. Soc. Am. **37**, 296 (1947).
[4] R. E. Richards and H. W. Thompson, Trans. Faraday Soc. **41**, 183 (1945).

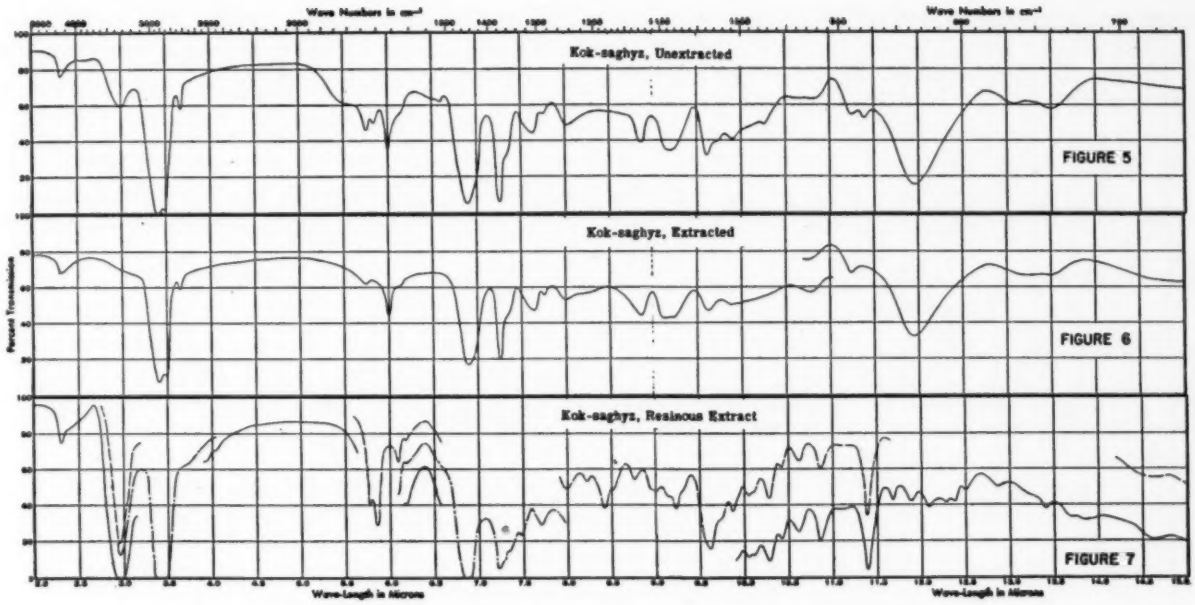


FIGURE 5

FIGURE 6

FIGURE 7

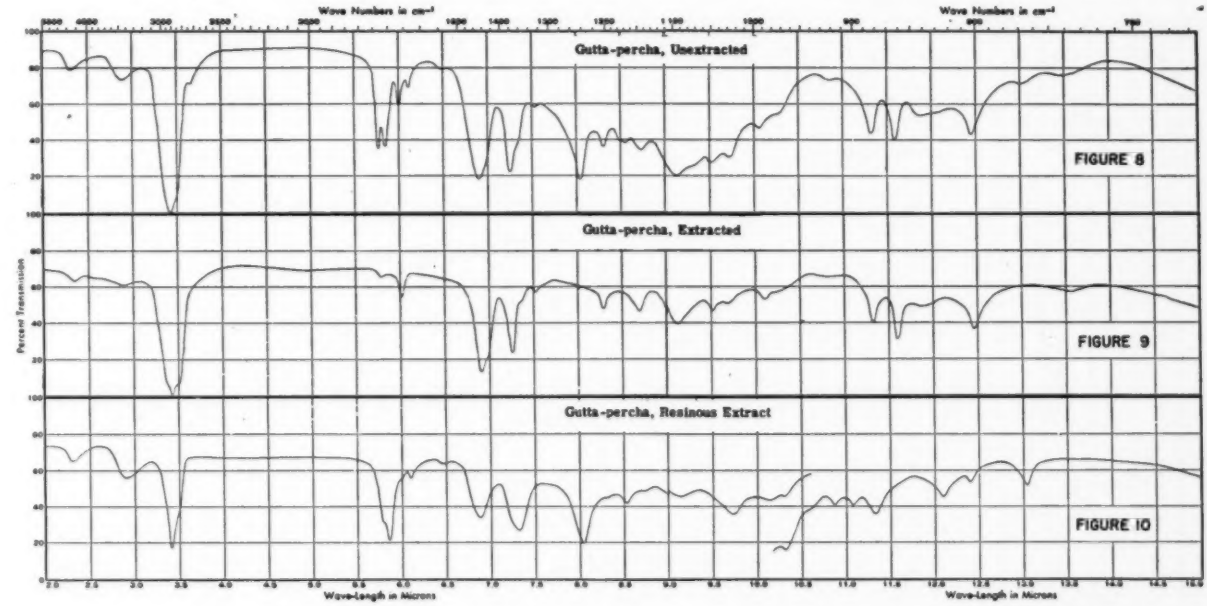


FIGURE 8

FIGURE 9

FIGURE 10

lence of molecular structure of the polymeric units in these three gutta-type elastomers.

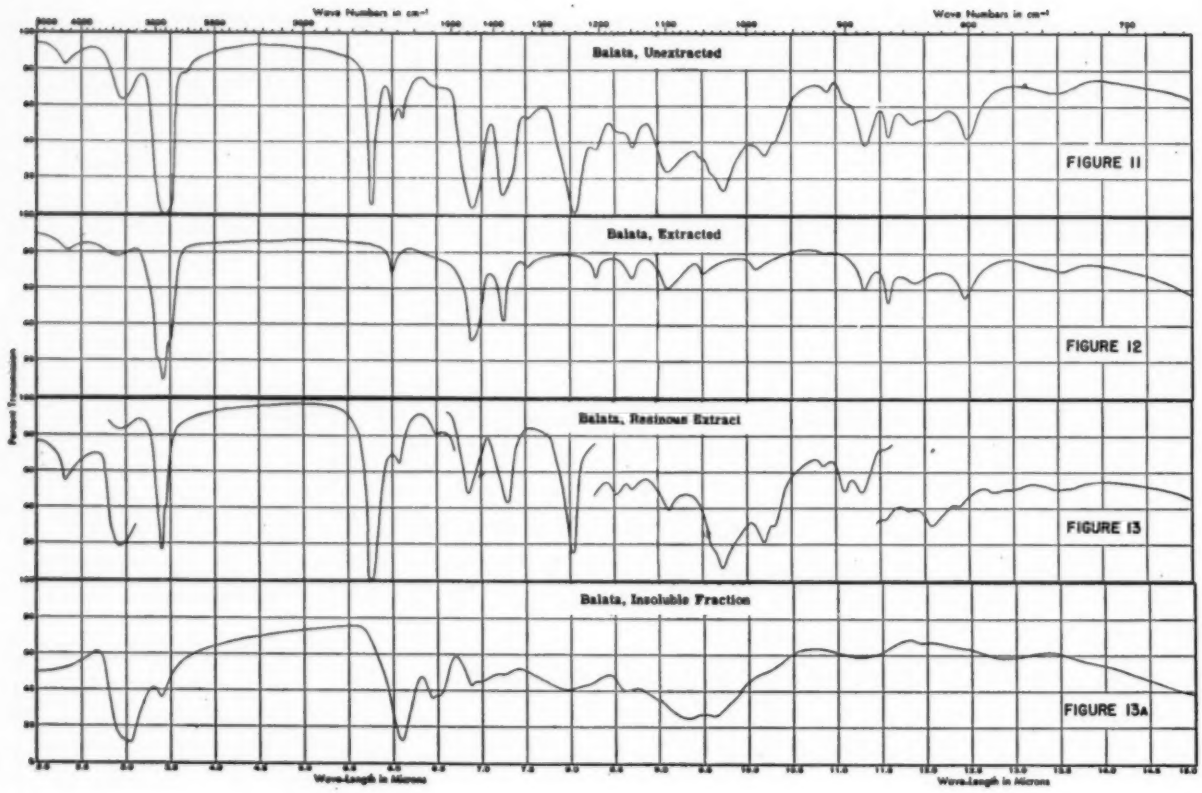
It may be noted that the infra-red spectrum, when measured at room temperature, differentiates sharply between hevea- and gutta-type elastomers. This is due, in part, to the fact that hevea-type elastomers are normally amorphous at room temperature whereas the gutta-type elastomers are in one of two possible crystalline modifications. Since the spectra of the crystalline modifications are quite different, as will be seen shortly, the identity of absorption of all gutta-type elastomers at room temperature as well as the differences between the spectra of hevea and gutta types depends upon the physical state of the samples. This dependence will now be discussed.

SPECTRAL EFFECTS OF CRYSTALLIZATION BY FREEZING: α- AND β-MODIFICATIONS OF GUTTA-TYPE ELASTOMERS

Hevea rubber, unlike the guttas, usually exists in the amorphous state at room temperature although upon long standing, i.e., for months or years, the rubber becomes hard or "frozen" because of crystal formation. At lower temperatures the rate of crystallization increases, and Wood[5] has shown that at −25°C the rate is maximum, half of the crystallization occurring in a few hours.

The spectra of hevea in the amorphous state (25°C) and in the crystalline state (−25°C) are shown in

[5] Wood, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. II.



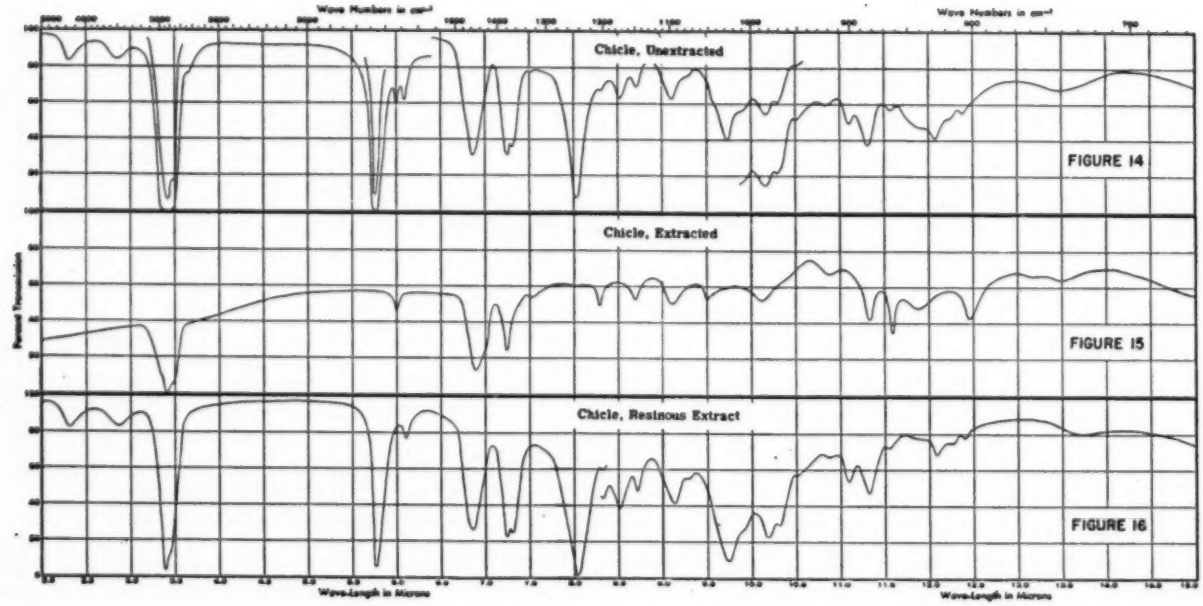


Fig. 17. In the latter case, the sample was held at −25°C for eight hours previous to and during the recording of the spectrum. The greatest spectral effect of freezing is the change in the position, shape, and intensity of the strong 11.95μ-band. A similar change has been observed previously by Williams and co-workers[6,7] for crystallization produced by stretching. A change in position and/or intensity is also noted for a few other bands, notably those at 8.8, 7.6, and 7.32μ; new weak bands for the crystalline state are observed at 11.49 and 10.38μ; the absorption around 8.05μ is resolved into three weak bands at 8.02, 8.12, and 8.26μ; and in general the bands tend to become sharp.

It is interesting to note that, as stated by Wood,[5] no crystallization occurs in samples of hevea having a molecular weight below 100,000. A sample of depro-

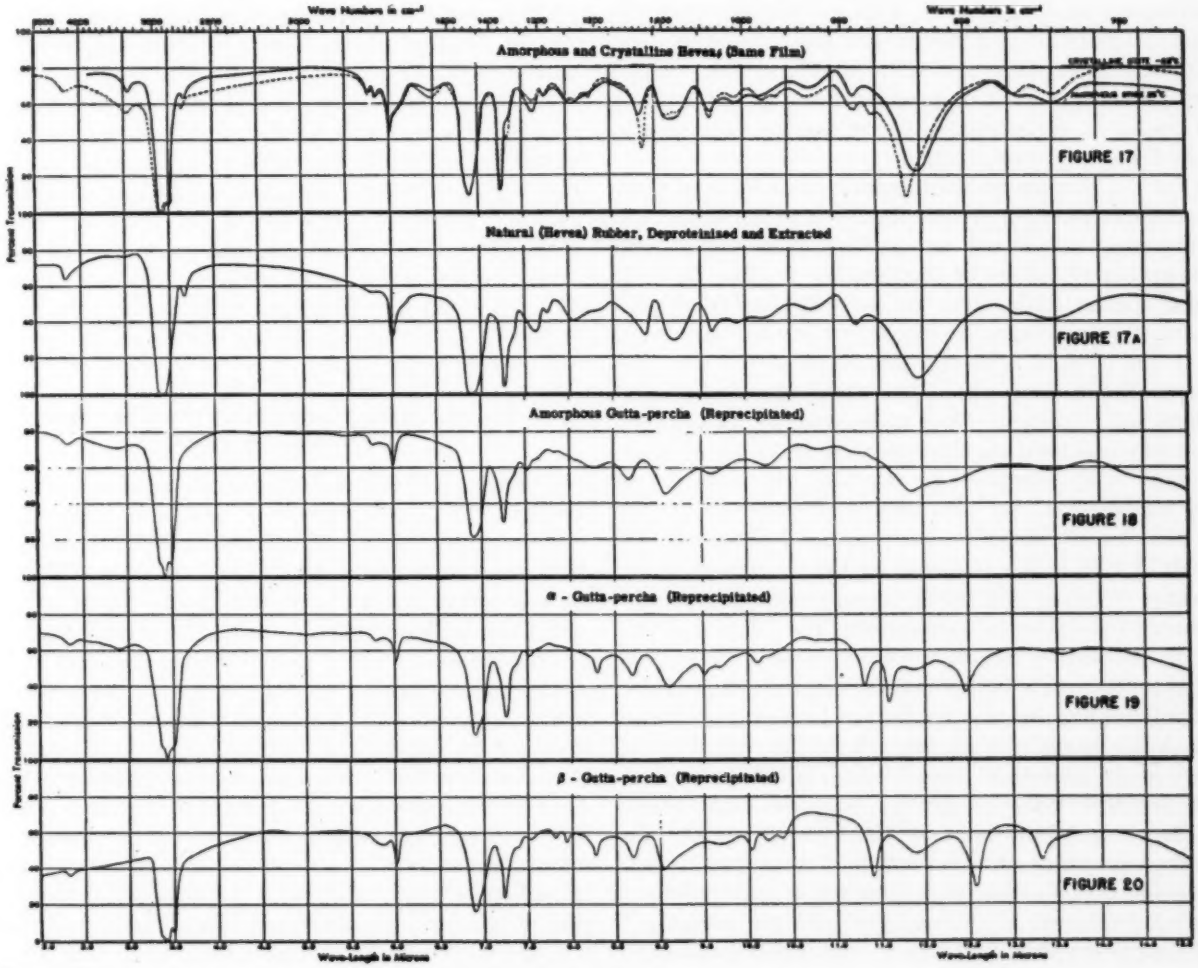
[6] D. Williams and R. Taschek, J. App. Phys. **8**, 497 (1937).
[7] D. Williams and B. Dale, J. App. Phys. **15**, 585 (1944).

teinized hevea was fractionated by precipitation from benzene solution with methyl alcohol and it was found that fractions having a viscosity average molecular weight below 100,000 showed no spectral evidence of crystallization after 8 hours at −25°C, while fractions above 100,000 showed crystallization after only two hours at this temperature. This phenomenon probably results from the greater thermal mobility of the shorter chains.[8]

It is well known that gutta-type elastomers can exist in two crystalline forms, α and β.[9] The spectra of the same film of purified gutta-percha in the amorphous state and in the α- and β-forms are shown in Figs. 18, 19, and 20, respectively. Spectra practically identical to these were also obtained for the α-, β-, and amorphous forms of balata and chicle. Films prepared by the evaporation of unheated solutions of the gutta-type elastomers in benzene are usually, but not always, in the α-form. Crystallinity is destroyed by heating the film to 80°C or higher, and when the film is allowed to cool by standing in air at room temperature, the β-form is nearly always obtained. It is possible to convert the β-form back to the α-form by heating to obtain the amorphous state and cooling slowly (0.5°C per hour).[5] Mixtures of the two crystalline forms are frequently obtained if the sample is not treated properly.

Crystallinity has a pronounced effect upon the absorption in the region of 11.5μ. In fact, the bands in this region may prove useful in studying phenomena related to crystal structure. The band at 11.59μ, found by Hendricks *et al.*,[10] as well as others at 11.32 and 12.46μ characterize the α-form, while those at 11.41, 12.57 and 13.32μ relate to the β-form. Other minor differences are observed in the region from 7 to 11μ. All previous measurements on gutta-type elastomers have been made at room temperature and the spectra therefore pertain to the crystalline forms.[10–12]

Now that the essential differences in the spectra of



[8] S. D. Gehman and J. E. Field, J. App. Phys. **15**, 371 (1944).
[9] C. C. Davis and J. T. Blake, *Chemistry and Technology of Rubber* (Reinhold Publishing Corporation, New York, 1937); also K. H. Meyer, *Natural and Synthetic High Polymers* (Interscience Publishers, Inc., New York, 1942).
[10] Hendricks *et al.*, Archives of Biochem. **7**, 427 (1945).
[11] Field, Woodford, and Gehman, J. App. Phys. **17**, 386–392 (1946).
[12] R. Stair and W. W. Coblentz, J. Research Nat. Bur. Stand. **15**, 295–316 (1935).

hevea- and gutta-type elastomers in their various physical states are known it is clear that spectral methods can be applied to practical problems of identification, analysis etc., of these elastomers with greater confidence than has heretofore been possible on the basis of previous experimental work. However, there remain a few important points in this connection which may now be cleared up. Foremost is the fact that all previous spectra of hevea, including that shown in Fig. 17, show evidence of impurities. Weak bands of variable intensity at 3.05 and near 6.50μ are usually present and several explanations for them have already been proposed. The clue to their origin, however, is found in the fact that proteinaceous materials in general, and particularly some which were isolated from gutta-type samples (see Fig. 13A) have strong bands at exactly these positions. This explanation is confirmed by the fact that the bands at 3.05 and 6.50μ are absent and absorption at 6.10 is somewhat weaker in the spectrum of a sample of commercially deproteinized hevea, shown in Fig. 17A. This is believed to be the first spectrum of pure hevea. For most purposes these impurity absorptions and a knowledge of their origins are unimportant, but in some instances, as in the study of oxidation, for example, it may become necessary to take them into consideration.

For the purpose of identification or analysis of these elastomers at room temperature the greatest spectral differences are found at positions associated with absorption of the α- or β-crystalline modifications. If these spectral differences are used for analytical purposes it is necessary, therefore, to insure beforehand that the sample is in the proper physical state. No preparational treatment of the sample has been found which will accomplish this with certainty and experience has indicated that the physical (crystalline) state is most easily identified by scanning the spectrum in the region of 11.5μ. The effects of crystallinity can be avoided, of course, by performing all measurements at 80°C, but at this temperature the differences in the spectra of hevea and gutta types are much less pronounced than at room temperature. The relation of this to the problem of structure will be discussed later. Another method of avoiding crystallinity effects is to use the intensity of the absorption at 11.9μ. At this wave-length the intensity of absorption, unlike that at other more likely analytical wave-lengths, is independent of the physical state of gutta-type elastomers (Figs. 18–20). It is significant that Hendricks chose the absorption intensity at this particular wave-length as a basis for classification of a number of naturally occurring elastomers according to *cis*- or *trans*-structure.[10] His conclusion that mixtures of hevea- and gutta-type elastomers are not found in nature is undoubtedly correct.

SPECTRA OF STRUCTURALLY RELATED COMPOUNDS

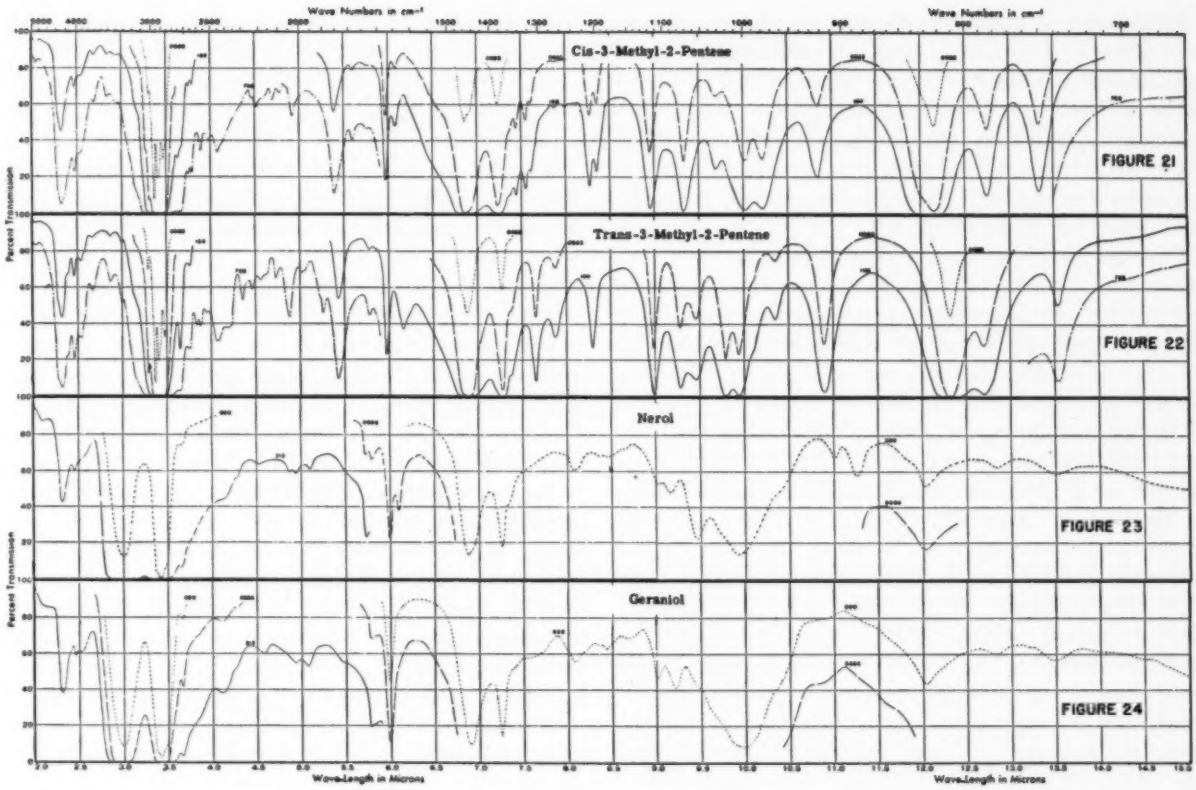
The somewhat academic but intriguing problem of identifying heavea- and gutta-type elastomers as *cis*- and *trans*-polyisoprene on the basis of their infra-red spectra alone has remained unsolved principally because of the lack of spectral correlations based upon related compounds of *cis*- and *trans*-structure. In an attempt to develop such correlations, the spectra of the isomeric alcohols Nerol (*cis*-) and Geraniol (*trans*-) and of the *cis*- and *trans*-isomers of 3-methyl-2-pentene have been studied. These compounds contain the polyisoprene structural unit and their spectra are shown in Figs. 21 to 28.

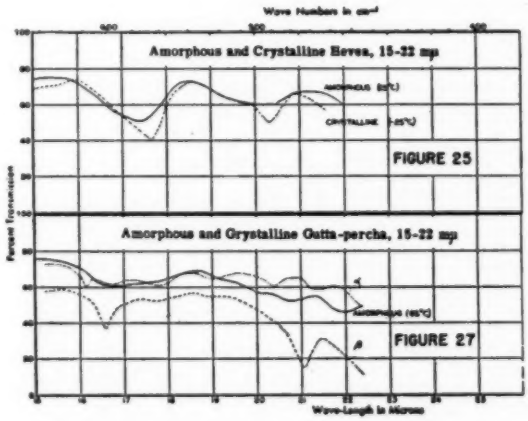
Since first members of a homologous series often exhibit anomalous spectral characteristics, the spectral differences observed for the isomeric hexenes would not be expected to correspond very closely to differences in the spectra of high molecular weight *cis-trans*-elastomers. However, the following similarities are noted. In the region of 6.0μ both Nerol and the *cis*-hexene have absorption at 6.10μ as does natural (hevea) rubber; the three *trans*-isomers have no absorption at this wave-length. Hevea and Nerol have a band at 11.25μ, while the *trans*-isomers, gutta and Geraniol, do not. Neither of the hexene isomers absorb at this wave-length. The obvious similarity at 5.7μ for the three sets of isomers has not been considered, since in the elastomers, bands in this region may arise from oxidation impurities. Hevea usually has two bands in this region, but both have not always been observed, as for example, in the case of deproteinized rubber (Fig. 17A). The correlations are consistent with the *cis*-configuration for hevea and the *trans*-configuration for gutta, but the question is by no means settled. It will be shown next, however, that polarization data furnish conclusive results.

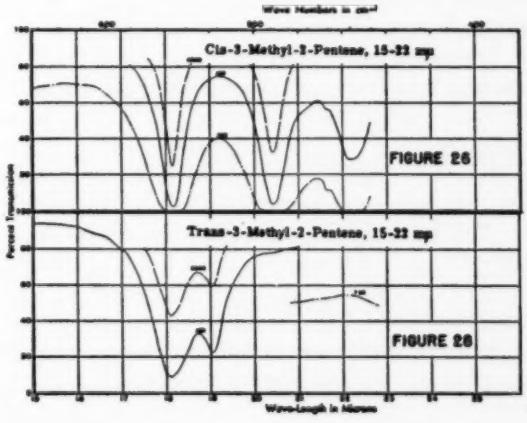
POLARIZATION SPECTRA OF ORIENTED CRYSTALLINE FILMS

In order to obtain additional information concerning the structural configuration of hevea- and gutta-type elastomers the spectra of hevea and gutta-percha have been redetermined using polarized radiation. This approach makes use of the dichroic properties of crystals which can be determined by measuring the transmission of oriented samples with plane-polarized incident radiation.[13] Crystallization in both types of elastomers is easily achieved and a preferred orientation of crystallites can be obtained by stretching or cold-rolling as has been shown by x-ray patterns. In an oriented sample some of the molecular vibrations will involve a changing dipole moment with axis parallel to the direction of orientation while others will be perpendicular or at some intermediate angle relative to the direction of orientation. When plane-polarized radiation is passed through the sample the absorption associated with each vibration will depend upon the component of the dipole moment which lies in the plane of polarization and perpendicular to the direction of propagation. In gen-

[13] See, for example, Elliott, Ambrose, and Temple, J. Chem. Phys. 16, 877 (1948).







eral the intensity of the bands will not be the same, therefore, when measured with radiation polarized parallel and perpendicular, respectively, to the direction of orientation. In this manner polarization measurements can provide useful information about the orientation of the various chemical bonds, provided that the nature of the vibrations associated with each absorption is known. Conversely, if the structure can be determined independently, polarization measurements can be used to great advantage in the spectral interpretation.

A polarizer constructed of silver chloride sheets, as described by Wright,[14] was mounted in the radiation path between the source and the sample (or entrance slit) in such a manner that it could be rotated through an angle of 180°. With this arrangement it was unnecessary to rotate the rubber specimen and errors due to inhomogeneity of sample film were thereby avoided.

A vulcanized sheet of natural (hevea) latex was used without purification for the present measurements since any impurity absorptions could readily be identified by comparison with the spectrum of pure hevea (Fig. 17A). Oriented crystals were produced by stretching unsupported films to several hundred percent elongation.

[14] Norman Wright, "Absorption spectra with polarized infra-red radiation," J. Opt. Soc. Am. 38, 663A (1948). See also R. Newman and R. S. Halford, Rev. Sci. Inst. 19, 270 (1948); Ambrose, Elliott, and Temple, J. Opt. Soc. Am. 38, 212–216 (1948). A different type of polarizer has been described by A. Elliott and E. J. Ambrose, Nature 159, 641–642 (1947).

In the case of gutta-percha, freshly purified samples were used and separate films of unoriented α- and β-crystalline modifications were prepared by the same methods used previously. These films were then stretched or cold-rolled to obtain orientation. Because of the poor tear resistance of gutta greater elongation and correspondingly higher orientation can generally be obtained by cold-rolling.

The spectra obtained are shown in Figs. 29–31. In the case of the hevea sample the spectral effects of crystallinity are somewhat less pronounced than those observed previously for crystallization produced by freezing. This suggests that more complete crystallization and perhaps better orientation might be obtained by freezing, followed by cold-rolling, rather than by stretching. However, a fairly high degree of orientation was evidently obtained, as indicated by the polarization of the band near 8.9μ so that the results obtained by the more convenient stretching method are undoubtedly satisfactory for the present purpose. The polarization data are in good agreement with earlier measurements by Hyde.[15] Gutta films in the β-form could be oriented without any apparent change in the extent or kind of crystallization but films which contained only α-crystallites before stretching invariably appeared to contain a small amount of β-crystallites after elongation, as indicated by the characteristic absorption bands in the region of 11.5μ. No attempt was made in this investigation to determine whether the appearance of β-crystallites upon elongation is caused by conversion of α into β or to additional crystal formation. Polarization spectra of gutta-type elastomers have not been reported previously.

MOLECULAR STRUCTURE OF HEVEA- AND GUTTA-TYPE ELASTOMERS

The first step in the interpretation of the experimental data is to choose a molecular model and see how well it accounts for the observations. Choice of the polyisoprene structure is dictated by a considerable amount of physical and chemical data and the simple, "ideal" structures for *cis*- and *trans*-forms of polyisoprene shown in Fig. 32 may be taken on an initial basis as the probable molecular configurations involved in hevea and gutta elastomers. If it is then assumed, as is customary, that the elastomer molecules are in the extended form in the crystallites and that the molecular axis is parallel to the direction of stretch in an oriented sample, definite predictions of the polarization of some of the well-known infra-red absorption bands can readily be made. In the *cis*-form of polyisoprene the double bonds are parallel to the molecular axis while in the *trans*-form they are inclined with respect to the molecular axis, as shown in Fig. 32. Therefore, if perfect and complete orientation of the crystallites could be achieved the 6.0μ-absorption band, which originates as a stretching of the double bond and involves a changing dipole moment lying essentially along the bond, would be nearly completely parallel-polarized in the *cis*-polymers and somewhat less completely in the *trans*-. Furthermore, if the molecules of both the *cis*- and *trans*-structures are twisted out of the planar configuration by any

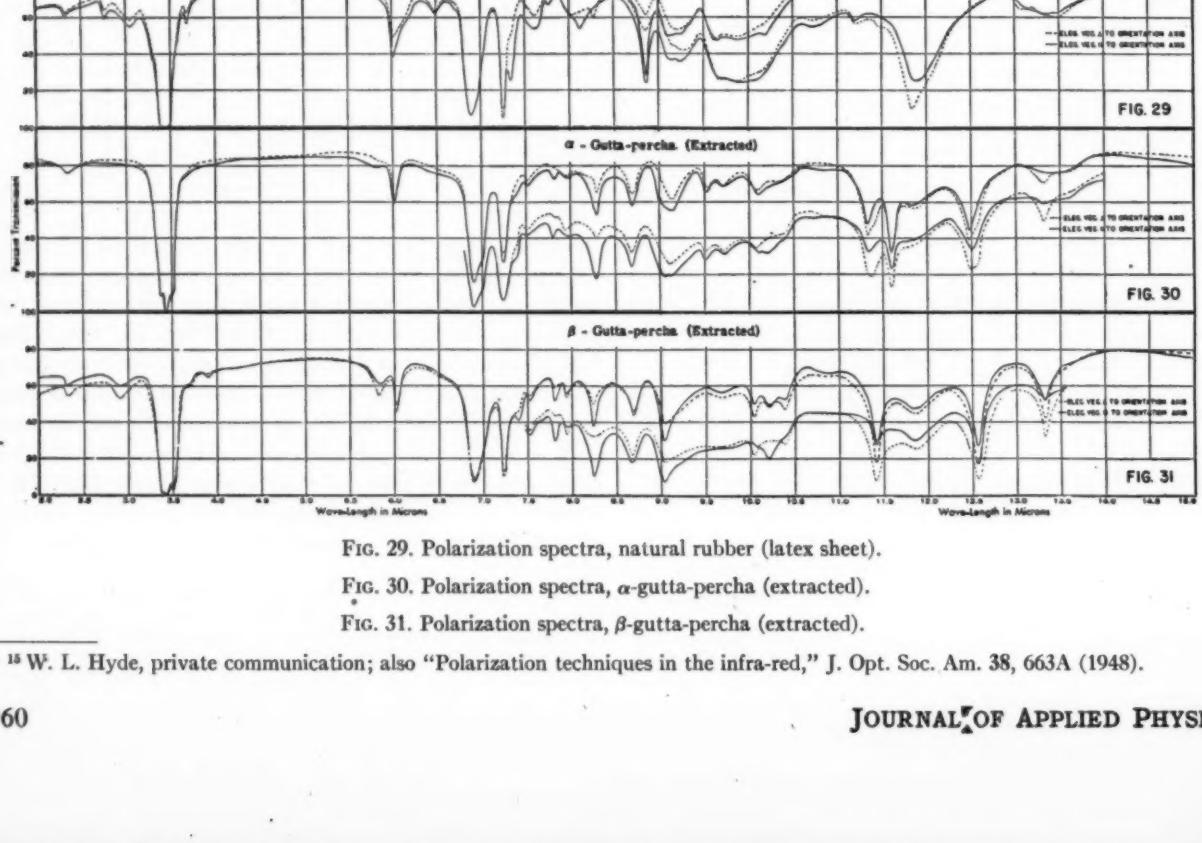


FIG. 29. Polarization spectra, natural rubber (latex sheet).

FIG. 30. Polarization spectra, α-gutta-percha (extracted).

FIG. 31. Polarization spectra, β-gutta-percha (extracted).

[15] W. L. Hyde, private communication; also "Polarization techniques in the infra-red," J. Opt. Soc. Am. 38, 663A (1948).

small amount the change in orientation of the double bonds relative to the molecular axis is such that the polarizations would be reduced but maintained in the same relative order. Since neither complete crystallization nor perfect orientation is obtained by stretching, the completeness of the observed polarization may be expected to be somewhat less than that which actually exists within the crystallites.

Reproducible measurements on several samples of each type of elastomer has shown that the 6.0μ-band in hevea is polarized parallel whereas in both the α- and β-forms of gutta it is less parallel-polarized and, in fact, appears on the average to be unpolarized or very slightly perpendicular (Figs. 29–31). These results cannot be due to incomplete orientation of the sample since other bands in the spectrum, e.g., that at 8.9μ, are almost completely polarized. The infra-red data therefore immediately lead to the assignment of the *cis*-configuration to hevea-type rubber and of the *trans*-configuration to gutta-type rubber in much the same manner as determination of the repeat distance by x-ray diffraction measurements has led to the same results using the same simple molecular models. Although the same basic assumptions are used in each case, it is significant that identical conclusions are reached by two entirely independent methods. In each case, however, the simple models shown in Fig. 32 are not completely satisfactory: in order to account for the x-ray patterns the molecules in the crystallites of both the *cis*- and *trans*-elastomers must be distorted relative to the planar configuration, as has been discussed recently by Bunn;[16] likewise, the polarizations of the infra-red band at 6.0μ show that the double bonds are not as nearly parallel to the stretch axis (molecular axis) as is indicated in the simple planar configurations of Fig. 32. While it is undoubtedly true that the change in dipole moment associated with the 6.0μ-absorption has components both parallel and perpendicular to the double bond and that this may account for the fact that the observed polarizations are not as nearly parallel as expected, it is nevertheless believed that the perpendicular component is small in the case of the isoprene structure and that the evidence actually favors a twisted rather than a planar configuration.

In order to develop additional information about the detailed configuration of the molecules in the crystallites it is first necessary to know something about the nature of the vibrations which give rise to some of the other absorption bands. Although several of the bands are known to be related to certain groupings[1] the vibrational modes are unknown except for the 6.0μ-band discussed above and some of the C—H vibrations. It seems more feasible, therefore, to attempt a spectral interpretation based upon polyisoprene models and using the polarization data as a guide. Although the planar models have already been shown to be unsatisfactory these models will be used for the assignment, assuming that the actual configurations correspond to similar but distorted models which, if available, would lead to the same assignment and to only slightly different polarizations.

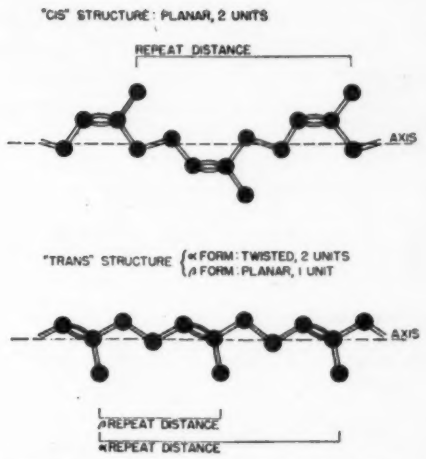


FIG. 32. Simple planar configurations for "*cis*-" and "*trans*-" polyisoprene.

TENTATIVE INTERPRETATION OF ELASTOMER SPECTRA

The polyisoprene unit, consisting of 13 atoms, will have 33 normal modes of vibration, each giving rise to an absorption band in the infra-red spectrum. As shown in the first column of Table 1, the vibrational motions can be described approximately in terms of twelve skeletal vibrations, six vibrations for each methylene group and nine vibrations of the methyl group. Since the isoprene units are not free, but joined together, there will be six additional motions which are listed at the end of Table I. The second column of Table I gives the spectral region in which each of the vibrations may be expected to occur, based upon a general (and incomplete) knowledge of the vibrations of hydrocarbons. All of the bending motions except those involving C—H bonds will be below 650 cm^{-1} and will not be found in the spectral region studied here, while the valency vibrations will all be above 650 cm^{-1} and within the regions studied. In columns three and four the approximate angle between the direction of change in dipole moment and the molecular (stretch) axis is given for the simple *cis*- and *trans*-configurations of polyisoprene shown in Fig. 32. A value of zero degrees indicates complete parallel polarization, 90 degrees indicates complete perpendicular polarization, and a value of 45 degrees indicates no polarization. For some vibrations an angular range is given, which depends upon the angle from which the molecule is viewed, the higher value being the more important in each case.

In addition to the help provided by the correlation rules and polarization data in interpreting the spectra of these elastomers there is one qualitative aspect pro-

[16] C. W. Bunn, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. II (Rubber), pp. 95–143.

TABLE I. Vibrational motions for polyisoprene.

Vibrational motions	Spectral region	Polarization* *Cis-*	*Trans-*	Hevea	Probable assignment α-gutta	β-gutta
Skeletal Modes (12)	cm^{-1}	Degrees		cm^{-1}	cm^{-1}	cm^{-1}
Stretching (5)						
C−H	~3050	55–0	80–0	3040	3060	3060
C=C	~1650	0	25–0	1665	1665	1665
C−CH_3	880–1090	55–0	0–0	980	980	980
C−CH_2 / C−CH_2	880–1200	55–0	30–0	1010–1090	~1095	~1110
Out-of-plane Bending (3)						
C−H Wag	800–850	90	90	845	~800	795
C−CH_3 Wag	<650	90	90	—	—	—
C=C Torsion	<650	90	90	—	—	—
In-plane Bending (4)						
C−H Bend	1260–1400	35–0	10–0	1285	1280	1280
C−CH_3 Bend	<650	35–0	10–0	—	—	—
C−C−C Angle	<650	0	25–0	—	—	—
C=C−C Angle	<650	28–0	85–0	—	—	—
Methylene Modes (6)×2						
Sym. Stretch	~3000	70–0	90	~2865	2855	2855
Unsym. Stretch	~3000	70–0	90	2930	~2925	~2925
Deformation	1440–1470	90	90	~1450	~1450	~1450
Rock	720–750	90	90	740, 765	750	750
Wag	~1300?	20–0	0	1130, 1350	1150, 1350	1150, 1350
Twist	~1250?	20–0	0	1230, 1315	1205, 1330	1215, 1330
Methyl Modes (9)						
Sym. Stretch	~2880	55–0	80–0	?	~2880	?
Unsym. Stretch (2)	~2950	55–0	80–0	2970	2980	2970
Sym. Deformation	1380	55–0	80–0	1380	1380	1380
Unsym. Deformation (2)	1400–1450	55–0	80–0	1450	1450(1430?)	1450
Rock	950–1150	35–0	20–0	1040	1050, 1030	1035
Wag	750–960	90	90	870, 890	865, 885	875
Torsion	?	—	—	—	—	—
Additional Modes (6)						
CH_2−CH_2 Stretch	700–1000	15–0	40–0	~840	~840	~840
Bendings (5)	<650	—	—	—	—	—

* Angle between vibrating dipole axis and molecular axis. Approximate values taken from Fig. 32. 90, 45, and 0° correspond to complete perpendicular polarization, no polarization, and complete parallel polarization, respectively.

vided by a consideration of the arrangement of the molecules in the crystal structure. Since the repeat distance in the β-form of gutta corresponds to one isoprene unit, the number of fundamental bands in the spectrum should correspond to the number of vibrations given in Table I, except that there may be two sets of CH_2 frequencies because of the different environment of the two CH_2 groups in the isoprene unit. But in hevea and in the α-form of gutta polymers the repeat distance corresponds to two isoprene units which would lead to the possibility of all the normal frequencies splitting into two components. Actually, some of the bands in the spectrum of β-gutta do appear to be split in the spectra of hevea and α-gutta, others appear to be broadened probably as a result of slight splitting, and in general the spectrum of β-gutta is the simplest. The experimental data is thus in qualitative agreement with expectations.

Some of the frequencies can be identified immediately on the basis of the correlation rules. Thus, the double-bond stretching and all of the C−H valency vibrations can be picked out, except in cases where the bands are not discreet because of lack in resolution. The frequency values in the region of 3000 cm^{-1} (3 to 4μ) were obtained from measurements with a lithium fluoride prism (as shown in Fig. 33). Most of the C−H bending vibrations can also be identified in a similar manner. The remaining vibrations are not as easily identified by the correlation rules and the help of the polarization data has been used in making a tentative, approximate assignment as indicated in the last three columns of Table I. Since the observed polarizations of the 6.0μ-band did not agree quantitatively with those predicted on the basis of the simple models used, it is obvious that only rough agreement between the observed and predicted polarizations can be expected for the other bands.

For *cis*-polyisoprene there should be three vibrations in the 700 to 1100 cm^{-1} region which are completely polarized perpendicular. There are five bands in this region which show appreciable perpendicular polarization in hevea. Since the out-of-plane C−H wagging vibrations are known to produce strong bands in the spectra of olefins, the strong band at 845 cm^{-1} is as-

signed to this mode. On the basis of the correlation data, the lower pair of bands at 740 and 765 cm^{-1} may then be assigned to the out-of-plane rocking of the two CH_2 groups and the remaining pair to a splitting of the out-of-plane CH_3 wagging mode. Except for the 6.0μ-band the only other vibrations which should be highly polarized parallel are the CH_2 in-plane wagging and twisting modes. Since each of these vibrations could give rise to two or even four bands, the parallel-polarized bands at 1360, 1315, 1230 and 1130 cm^{-1} may correspond to these modes, although the last one may be the overtone of a band at 570 cm^{-1}. One reasonable but tentative interpretation of the remaining bands is to assign the unpolarized band at 1285 cm^{-1} to the C—H in-plane bending mode, 1040 cm^{-1} to the in-plane CH_3 rock, 980 cm^{-1} to the CH_3 stretch, and the remaining absorption in the region 1000 to 1100 cm^{-1} to the C—C (one carbon unsaturated) stretching modes. And lastly, the only other vibration which should appear above 650 cm^{-1} is a paraffinic C—C valency vibration which presumably occurs near 840 cm^{-1} and accounts for part of the intensity and asymmetry of the band in this region. Since the band at 1540 cm^{-1} in Fig. 29 is caused by impurities, the only bands which are unaccounted for by this assignment are those at 1250, 1210, and near 930 cm^{-1}, and those might easily be caused by combinations involving the several unknown bending modes at low frequencies.

In a similar manner reasonable assignments can also be made for the spectra of α- and β-gutta-percha (Table I). In both cases there is a band near 1260 cm^{-1} and weak absorption near 930 cm^{-1} which is unaccounted for, except as combinations involving lower frequencies.

DISCUSSION

One part of the assignments given in Table I, namely the vibrations of the methyl group, can be checked by studying the spectrum of Neoprene. According to x-ray measurements[16] the configuration of Neoprene is analogous to that of β-gutta-percha. In any case, the molecular structure of Neoprene is that of polyisoprene in which all methyl groups are replaced by chlorine. The polarization spectra were obtained (Fig. 34) and all of the bands assigned to vibrations of the methyl group are absent in accordance with expectations. The observed perpendicular polarization of the 6.0μ-band is also in agreement with the *trans*-configuration about the double bonds in the elastomer chain.

The methylene frequencies which have been assigned to bands in the region 1100 to 1360 cm^{-1} are not easily distinguished since the in-plane C—H bending frequency and probably some combination frequencies also fall in this region. Until more is learned about the correlation rules for these vibrations, or until additional data can be obtained, as, for example, by the study of deuterium substitution at specific positions, the interpretation of this spectral region cannot be made with confidence. It is quite possible for example, that the C—H bending and one of the methylene frequencies should be interchanged.

The assignments given in Table I appear to be quite reasonable, however, and since the interpretation was made on the basis of a polyisoprene model the infra-red

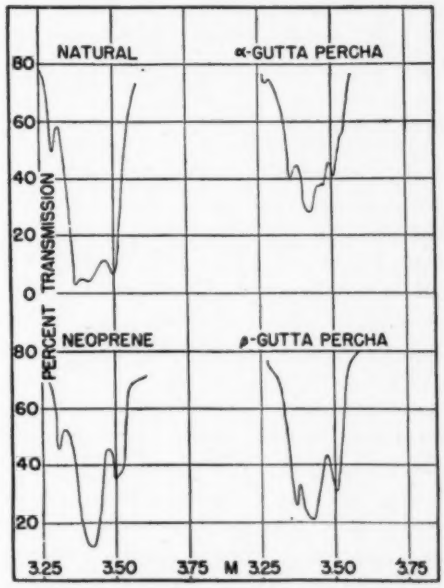


FIG. 33. Spectra of natural (hevea), α- and β-gutta-percha and neoprene in the 3.5μ-region, as obtained with lithium fluoride prism.

spectra may be considered as being in agreement with this type of molecular structure. If it is assumed that the assignments are correct it should then be possible, on the basis of the polarization data, to arrive at some conclusions regarding the relative orientation of some of the chemical bonds, i.e., the detailed structural configuration within the crystallites. This approach is not very desirable, however, since the polarizations have already been used in making the assignments. Furthermore, several other complicating factors arise: accurate measurements of the dichroic ratio are required, and in the case of elastomers this is difficult because of the effects of incomplete crystallization and imperfect orientation in the sample, polarizer inefficiency and errors in the measurements, effects of unresolved or overlapping bands, and because of the fact that for a given orientation of the molecular axis the polarization of bands associated with some vibrations depends upon the angle from which the molecule is viewed and therefore varies with the amount of two-dimensional orientation. Thus, it is probable that only the direction (parallel or perpendicular) and amount (strong or weak) of polarization has significance in the case of some bands. Bearing these facts in mind, the observed polarizations (Figs. 29–31) are nevertheless in rather poor agreement with the planar configurations (Fig. 32), the results for the 6.0μ-bands, as discussed earlier, being of particular importance. The configurations proposed by

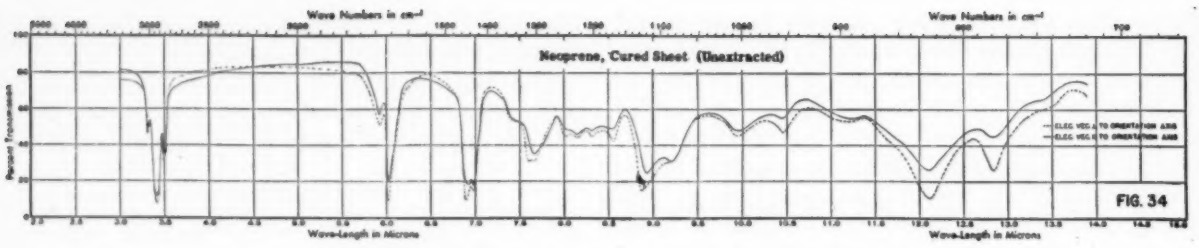


FIG. 34. Polarization spectra, Neoprene, cured sheet (unextracted).

Bunn,[16] however, are in fairly close agreement with the infra-red data, especially in the case of the β-form of gutta-percha where the model predicts very little if any polarization of the 6.0μ-band, in agreement with observation. In the case of the *cis*-rubber, however, the agreement is not as good. The infra-red data for example, indicate that the double bonds are inclined to the stretch axis by an amount somewhat greater than that predicted by the x-ray diffraction model. No model so far proposed is entirely satisfactory in accounting for all the physical and chemical data,[5] however, and in addition to the many other criteria any satisfactory model must account for the polarizations of the infra-red bands, particularly that at 6.0μ caused by the C=C stretching vibration. If such a model is found it may become necessary, in turn, to change some of the assignments proposed here in order to account for the observed polarizations of all other bands.

Finally, it is of interest to see whether the proposed assignments suggest any reasonable explanation for several spectral observations, e.g., the spectral changes occurring upon crystallization by freezing or stretching, and the spectral differences between hevea and gutta in the amorphous state.

In the case of hevea (see Fig. 17) the shift toward higher frequency and intensification of the 840 cm^{-1} band may now be associated with a modification of the forces governing the wagging mode of the hydrogen attached to the double bond. In the amorphous state greater thermal agitation of the molecules and variable but less influential environmental forces tend to broaden this band and make it less discreet than in the crystalline state. Similar effects are also active in the case of many other bands which show a shift in position and change in intensity upon crystallization, and in the case of the methylene vibrations occurring in the range 1100 to 1360 cm^{-1} an additional sharpening of the bands and splitting of some of them into discrete components is brought about by arrestment of rotation about the C−C bonds. The appearance of the new band at 870 cm^{-1} can be attributed to the presence of non-equivalent methyl groups in the crystallites, as discussed previously. The same explanation applies also to the new band at 963 cm^{-1}. This band was not observed in the case of the stretched sample, however, possibly as a result of interfering absorption of the compounding ingredients in this region. Otherwise the spectra of crystalline samples produced by freezing and by stretching are identical and the explanation of the spectral changes observed upon crystallization are undoubtedly the same in both cases.

The spectral changes which occur upon crystallization of gutta percha are considerably more pronounced than in the case of hevea elastomers. This is especially so in the region of 700 to 900 cm^{-1} where some of the strong bands found in the spectra of the crystalline samples are hardly discernable in that of the amorphous sample (see Fig. 18). It will be noted, however, that these bands have been assigned to molecular vibrations and on the basis of this assignment the change in absorption upon crystallization is simply due to an unusually great sharpening and probably also to an intensification of the very broad weak bands in the spectrum of the amorphous sample. These absorptions in the regions of 800 and 875 cm^{-1} have been assigned to the wagging modes of the hydrogen and methyl, respectively, and since both the hydrogen and the methyl form part of the rigid portion of the elastomer chain it is somewhat surprising that these bands are not more discrete and well defined in the amorphous state. The sharpening of the bands in the region of 1100 to 1350 cm^{-1} which have been assigned to vibrations of the methylene groups is to be expected, however, due to freezing of the molecular configuration upon crystallization.

The differences between the spectra of the α- and β-crystalline modifications of gutta are also of interest. These spectral differences must result from differences in the molecular configuration within the crystallites and it should be expected that they are associated with vibrations involving the methylene chain since this is the only part of the molecule which is free to assume different configurations. The marked spectral differences near 875 and 800 cm^{-1}, however, have been assigned to vibrations involving the rigid part of the molecule whereas the absorption in the region 1100 to 1350 cm^{-1}, which has been assigned to methylene vibrations, is nearly identical. An explanation of this is found in the fact that the methylene vibrations occurring in the spectral range studied here consist mainly of hydrogen motions and are therefore largely independent of the configuration of the carbon chain. It is rather the chain vibrations involving stretching and bending of the C−C bonds which should be dependent upon the

skeletal configurations. These valency vibrations have been assigned to the weak bands in the region 1000–1100 cm^{-1} where differences do occur, and to a band at 840 cm^{-1} where no difference occurs. The skeletal bending vibrations, however, are not found in the spectral region studied here but at longer wave-lengths and it is very probable that the greatest spectral differences between α and β actually occur in that region in accordance with expectations. That the spectral differences in the regions of 800 and 875 cm^{-1} (where the bands have been assigned to vibrations of the rigid part of the molecule) are as great as they are is apparently due to differences in the molecular environment of the rigid grouping in the two crystalline phases. This arises because of different molecular configurations and possibly because of different relative orientations of the molecules within the crystalline lattices.

The remaining point of interest, one analogous to that just discussed, involves the differences in the observed spectra of hevea and gutta in the amorphous state (Figs. 17A and 18). The observed differences are not unusually pronounced and are presumably due to *cis-trans*-isomerism. In this case the spectral differences should be expected to be associated with bands resulting from vibrations of the rigid part of the molecule rather than to vibrations of the non-rigid methylene chain. While differences are observed throughout the region 1350 to 700 cm^{-1} it will be noted that several of the greater differences are found in the 1350 to 1100 cm^{-1} region where the bands have been assigned to methylene vibrations and not to vibrations of the rigid configuration about the double bonds. It is not surprising, however, that the methylene vibrations are different since the molecular environment of the methylene hydrogens is not the same for the *cis-* and *trans*-configurations (see Fig. 32). That the differences in other spectral regions are not greater than those observed is probably caused by the great broadening of the bands which occurs in both spectra and which tends to obscure the spectral differences.

Although the spectral interpretations which have been presented here appear to be reasonable and useful in accounting for the experimental observations, additional efforts should be made to complete the interpretation and to determine the validity of the assignments. A complete interpretation of the elastomer spectra is of considerable importance to rubber research, for full utilization of infra-red methods can be achieved only with the help of such basic knowledge. For example, when an observable change in the spectral characteristics of the elastomer is produced upon treatment of the sample, as in the study of oxidation or other forms of degradation, a knowledge of the vibrational assignment may make it possible to interpret the spectral change and thus to deduce the point and nature of attack on the elastomer chain. Although the present interpretations are incomplete and tentative, it is suggested that they provide a basis for the initiation of such work.

ACKNOWLEDGMENTS

Acknowledgment is made to Mr. J. J. O'Niell who performed all the work on guayule included here. Thanks are also extended to Dr. Peter King of this Laboratory for his helpful advice throughout the work.

Magnetic Characteristics of an Oriented 50 Percent Nickel-Iron Alloy*

J. H. Crede and J. P. Martin
Research Laboratory, Allegheny Ludlum Steel Corporation, Brackenridge, Pennsylvania
(Received February 16, 1949)

Single crystal magnetic properties have been closely approached in a polycrystalline 50 percent nickel-iron by the development of a favorable grain orientation. The method is adapted to commercial production by careful control of processes effecting the fundamental magnetization process in the magnetic domain. Elimination of steps I and III of the normal magnetization has produced a nearly geometrically true rectangular shape of the hysteresis loop.

INTRODUCTION

NUMEROUS investigators[1] have found that the magnetic properties of single crystals depend upon the crystallographic direction in which these properties are measured. Figure 1 illustrates these variations for single crystals of iron and nickel. The 50 percent nickel-iron alloy presents an anomaly in that the easy direction is the same as for iron but the directions of intermediate and most difficult magnetization are reversed as compared to iron.

The use of single crystals in most commercial applications is obviously impractical and it became evident that if the orientation of crystals in a polycrystalline material could be controlled then the magnetic properties in one direction could be made to approach the best properties obtainable in single crystals. W. S. Smith *et al.*[2] and Goss[3] developed preferred crystal orientation in magnetic materials with moderate cold reductions while Dahl and his coworkers[4] and Snoek[5] produced a high degree of preferred crystal orientation in the nickel-iron alloys containing 40 to 90 percent nickel by severe cold reductions of the order of 98 percent. Both processes require a suitable heat treatment to develop the optimum magnetic characteristics. Dahl and Pawlek extended to the grain oriented nickel-iron alloys the investigations of Kelsall, Dillinger, and Bozorth[6] on the effect on magnetic properties of heat treatment in the presence of a magnetic field and obtained an additional degree of perfection in the achievement of single crystal properties.

The steep slope of the magnetization curve and correlated rectangular shape of the hysteresis loop are encountered in single crystals of ferromagnetic materials in the crystallographic direction of easiest magnetization. Similar characteristics have been observed in some polycrystalline magnetic materials having small values of the anisotropy constant when cooled in the presence of a magnetic field or subjected to a homogeneous mechanical stress. The grain oriented 50 percent nickel-iron alloy, although consisting of a multitude of fine grains or crystals, behaves magnetically nearly as a single crystal because of a majority of the individual grains are oriented in such a manner that a cube edge is parallel to the direction of rolling with a cube face in the plane of the strip. Since a cube edge is parallel to the direction of easy magnetization for 50 percent nickel-iron, a rectangular shape of the hysteresis loop is to be expected when the measuring field is applied in the direction of rolling.

DISCUSSION

The existence of directions of easiest magnetization and the correlated rectangular hysteresis loop can be logically explained on the basis of the domain theory of ferromagnetism.[7] According to theory, material in the ferromagnetic state is comprised of regions or domains in which the electron spins of the atoms are aligned by exchange forces in a direction requiring a minimum expenditure of energy. The domain is said

* A major portion of this work was done under contract W-36-039 SC32301 with the Signal Corps Engineering Laboratory, Fort Monmouth, New Jersey. The authors wish to thank this agency for permission to publish the results of this investigation.

[1] K. Beck, "International critical tables of numerical data," Physics, Chemistry and Technology, Vol. VI, pp. 410; W. L. Webster, "Magnetic properties of iron crystals," Proc. Roy. Soc. London **A107**, 496–509 (1925). K. Honda, and S. Kaya, "On the magnetization of single crystals of iron," Sci. Report, Sendai, series 1, **15**, 721–753 (1926); K. J. Sixtus, "Magnetic anisotropy in silicon steel," Physics **6**, No. 3, 105–111 (1935); J. D. Kleis, "Ferromagnetic anisotropy of nickel-iron crystals at various temperatures," Phys. Rev. **16**, 39–62 (1936); F. Lichtenberger, "Magnetostriction and magnetization of single crystal iron-nickel alloys," Ann. d. Physik, **15**, 45–71 (1932); R. M. Bozorth, "Directional ferromagnetic properties of metals," J. App. Phys. **8**, 575–588 (1937); W. E. Ruder, "Magnetization and crystal orientation," Trans. Am. Soc. for Steel Treating **8**, 23–29 (1925).

[2] Smith, Garnett, and Randall, U. S. Patent 1,915,766.

[3] N. P. Goss, "New development in electrical strip steels characterized by fine grain structure approaching the properties of a single crystal," Transactions of the American Society for Metals, **23**, 511–544 (1935).

[4] O. Dahl and J. Pfaffenberger, Zeits. f. Physik, **71**, 93–105 (1931); also Metallwirtschaft,—wissenschaft und—technik, **14**, 25–28 (1935); O. Dahl, and F. Pawlek, Zeits. f. Physik, **94**, 504 (1935); F. Pawlek, "Effect of fiber structure and of cooling in a magnetic field on magnetization," Zeits. f. Metallkunde, **27**, 160–165 (1935).

[5] J. L. Snoek, Physica, **2** (4), 403–412 (1935); W. G. Burgess, and J. L. Snoek, "Structures of rolled and of recrystallized iron-nickel-alloys," Zeits. f. Metallkunde, **27**, 158–160 (1935).

[6] G. A. Kelsall, "Permeability changes in ferromagnetic materials heat treated in magnetic fields," Physics **5**, 169–172 (1934); J. F. Dillinger and R. M. Bozorth, "Heat treatment of magnetic materials in a magnetic field, Part I," *ibid.* **6**, 279–284 (1935); R. M. Bozorth and J. F. Dillinger, "Heat treatment of magnetic materials in a magnetic field, Part II," *ibid.* **6**, 285–291 (1935).

[7] R. M. Bozorth, "The physical basis of ferromagnetism," Bell Sys. Tech. J. **19**, 1–39 (1940); R. M. Bozorth, "Magnetism," Rev. Mod. Phys. **19**, 29–86 (1947).

to be saturated because the atomic magnets are all pointing in the same direction. In the unmagnetized state, the individual domains are arranged randomly in all the easy directions so that the net magnetization of the specimen as a whole is zero. When a weak magnetic field is applied those domains favorably oriented with respect to the applied field will increase in volume at the expense of neighboring domains less favorably oriented. This change in volume by reversible domain boundary displacements imparts a resultant magnetic moment to the specimen. As the field is gradually increased reversible domain boundary displacements become irreversible with abrupt increases of the induced magnetism. This latter phenomenon is the well-known Barkhausen effect. The field strength required for the Barkhausen effect is somewhat lower than the coercive force and usually it is clearly distinguished from the initial field region, where the displacement process predominates, by the sharp rise of the magnetization curve. With a further increase in the field the magnetization vectors are rotated away from the easy direction nearest the direction of the applied field and are forced to align themselves in the direction of the field, at which point the specimen attains magnetic saturation.

The magnetization process in a grain-oriented material parallel to an easy direction differs from that of a randomly oriented material or a grain-oriented material in which the field is not parallel to an easy direction and is manifested in the rectangular shape of the hysteresis loop. Reversible domain growth and rotation processes accounting for about 50 percent of the total magnetization in randomly oriented specimens are almost entirely eliminated in a specimen having a high degree of preferred orientation. In grain-oriented materials the preponderance of 180° reversals which take place with a minimum energy loss reduce the potential gradient between adjacent domains for the promotion of the growth of one domain at the expense of another. Due to the equivalence of crystallographic directions of easy magnetization other than the easy direction parallel to the field, there remains a certain amount of 90° jumps contributing to the total magnetization. It is possible that some domain growth may occur since the 90° jumps require a definite expenditure of energy. With the elimination of the 90° jumps by annealing in a magnetic field, an almost geometrically true rectangular shape of hysteresis loop can be obtained in some materials.

The stresses caused by magnetostriction are of particular importance since stresses will tend to change the direction of easy magnetization of the domains more nearly parallel or normal to the direction of the stress accordingly as the magnetostriction (the change in length that occurs when a ferromagnetic body is magnetized) is positive or negative. The alignment of domains in an applied field by magnetostrictive stresses can be accomplished most easily in materials of low anisotropy in which domain orientation is only weakly controlled by crystal structure. This process is facilitated by increasing the temperature to a point just below the Curie temperature so that plastic flow may occur allowing the domains to change their shape and so relieve the stresses of magnetostriction. The vector of domain magnetization after cooling will stay in the position which it assumed during the anneal when the crystal anisotropy is low. In materials of high crystal anisotropy it will tend to return to the closest direction

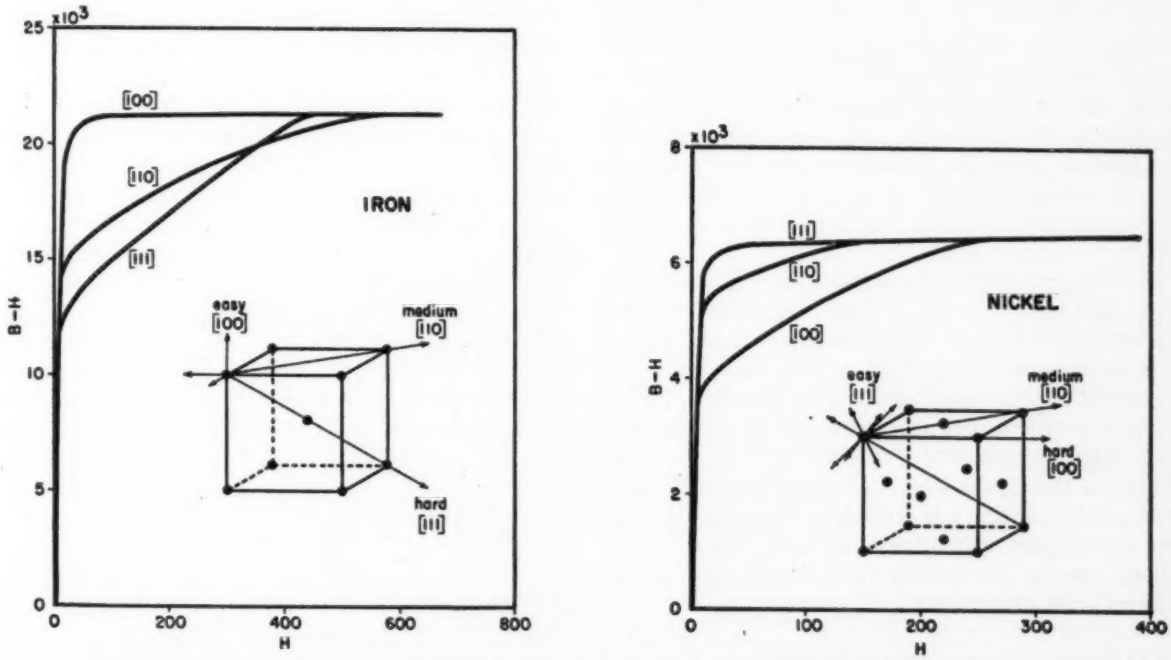


FIG. 1. Magnetic properties and crystal structures of single crystals of iron and nickel.

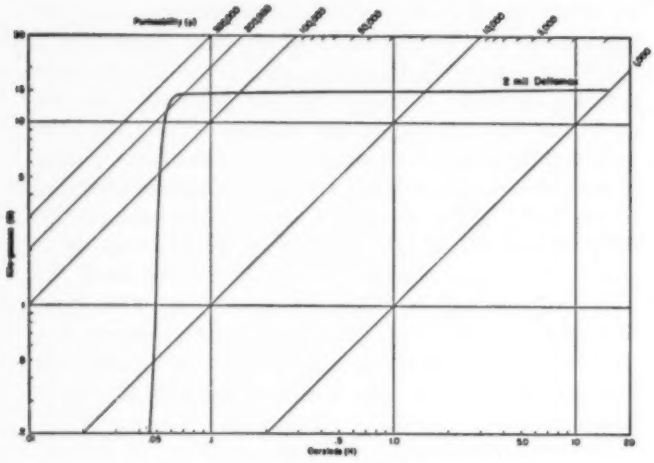


FIG. 2a. D.c. magnetization curve.

of easy magnetization so that the crystal forces will balance the magnetostrictive stresses. After demagnetizing, subsequent magnetization will proceed mainly by domain reversals, thereby extending the steep part of the magnetization curve at low field strength up almost to the saturation value. In the case of materials with high crystal anisotropy the final rotation of the magnetization vector into the direction of the field to achieve complete saturation is then aided by the magnetostriction. In either case a rectangular shape of the hysteresis loop is achieved except for the small deviations from ideal conditions for the alignment. In the grain-oriented 50 percent nickel-iron alloy, however, the direction of the field during annealing can be made to coincide with an easy direction of magnetization. In this case the magnetic anneal will eliminate almost completely the 90° jumps which still contribute appreciably to the magnetization process in a material which is merely grain oriented. When the material is demagnetized at room temperature half of the domains will undergo 180° reversals. If it is now magnetized in the preferred direction complete saturation will be achieved by means of 180° reversals and a geometrically true rectangular shape of the hysteresis loop should result.

Controlling or eliminating impurities and mechanical stresses and the perfection of grain orientation are the most important factors in determining the quality of the grain-oriented 50 percent nickel-iron alloy. The detrimental effect of small amounts of impurities on the magnetic quality of iron and nickel-iron alloys has been convincingly described by Yensen and Cioffi.[8,9] In addition to the deleterious effect on permeability and coercive force, increasing amounts of impurities will lower the saturation induction and alter the recrystallization process preventing the attainment of an optimum grain orientation.

[8] T. D. Yensen, Phys. Rev. **39**, 358–363 (1932).
[9] P. P. Cioffi, Phys. Rev. **39**, 363–367 (1932).

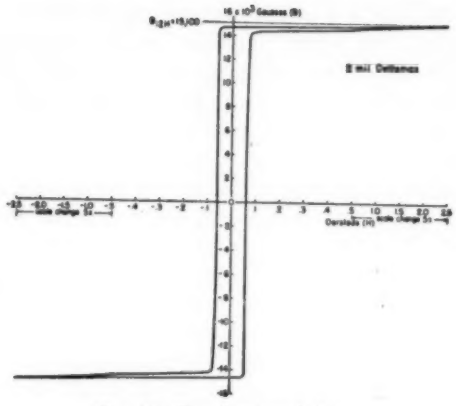


FIG. 2b. D.c. hysteresis loop.

The development of the mechanical rectifier in Germany by Koppelmann[10] during the early years of World War II created a demand for the commercial development of a material having rectangular hysteresis loop characteristics. The choke coils for the mechanical rectifier require a magnetic material having a rectangular hysteresis loop with sharply expressed knees, a high ratio of retentivity to the saturation value and a low value of coercive force. The delta-*B* value (useful magnetization range) should be in excess of 26 kilogausses. Although rectangular hysteresis loops were known to exist in the 68 percent nickel Permalloy and Perminvar (45 percent nickel, 25 percent cobalt, 30 percent iron) when subjected to cooling in a magnetic field, it was apparent to the German scientists that in order to fulfill the exacting requirements of the core material for the mechanical rectifier, it was most desirable to perfect a process for the orientation of the 50 percent nickel-iron alloy because of the higher intrinsic saturation and consequently higher delta-*B* value to be expected. A method of manufacturing a grain-oriented 50 percent nickel-iron alloy in light gauge tape form, known as Permenorm 5000-Z, was developed in Germany by Heraeus-Vacuumschmelze employing the 98 percent cold reduction described by Dahl and Snoek. A discussion of the processing methods involved in the production of this material and its subsequent fabrication into toroidal cores was described in Technical Memorandum Number M-1091[11] issued by the Signal Corps Engineering Laboratories. A supplementary article, Technical Memorandum Number M-1137[12] describing the requirements necessary for the achievement of rectangular hysteresis loop materials, was presented at the Magnetic Materials Symposium[13]

[10] F. Koppelmann, Elektrotech. Zeits. **62**, 3–20 (1941).
[11] E. Both, "Plant scale process for the production of rectangular hysteresis loop magnetic materials," Signal Corps Eng. Lab., Tech. Memo. No. M-1091 (September, 1947).
[12] E. Both, "Some factors governing the achievement of a rectangular hysteresis loop, especially in grain oriented 50 percent nickel-iron," Signal Corps Eng. Lab., Tech. Memo. No. M-1137 (June, 1948).
[13] Papers presented at the Naval Ordnance Laboratory Magnetic Materials Symposium, U. S. Naval Ordnance Publication (June 15, 1948).

sponsored by the Naval Ordnance Laboratory on June 15, 1948 in Washington, D. C. The magnetic characteristics of Permenorm 5000-Z were made public at this symposium and an immediate interest was aroused regarding new commercial applications.

The German technique for the commercial production of Permenorm 5000-Z entailed the most careful selection of high purity raw materials and a melting practice conducted under reduced pressure or in vacuum. Additional purification was obtained during various intermediate stages of processing the ingot into strip by annealing in an atmosphere of purified dry hydrogen. According to German views, the practice of slitting as a final operation in the required thicknesses of one and two mils was undesirable because of the detrimental effect on the magnetic characteristics. Consequently, material was slit to within one percent of the final width at an intermediate thickness of 14 mils and processed to a thickness of two mils by continued cold reduction on a specially designed six stand, six high tandem mill. Further reduction in gauge was accomplished on a small twenty-roll cluster type Rohn mill.

EXPERIMENTAL

The Allegheny Ludlum Steel Corporation for some time prior to public release of information on Permenorm 5000-Z had been requested by the Signal Corps Engineering Laboratories to investigate the possibility of producing on a commercial basis material having similar magnetic characteristics. The first phase of this study was devoted to perfecting the technique of winding, insulating, and annealing wound cores having the desired rectangular shape of the hysteresis loop. This research was conducted on material of German manufacture supplied by the Naval Ordnance Laboratory. After this phase of the investigation was successfully completed the more difficult task of developing a method for the commercial production of a material of similar characteristics was initiated.

It is obvious that the highly specialized facilities developed in Germany for melting under reduced pressure and rolling to extremely light gauges in tape form would impose severe restrictions on commercial production in this country. Recognizing this fact, efforts were directed along lines which would permit utilizing conventional melting and rolling facilities for the production of a material of comparable magnetic quality. These characteristics are exemplified in Allegheny Deltamax, a product heretofore unavailable to American industry; a material possessing rectangular hysteresis loop characteristics; exceptionally high maximum permeability; and low coercive force in a high saturation grain oriented 50 percent nickel-iron alloy. It has been found feasible to melt and refine this alloy to an exceptionally high degree of purity by the careful selection of raw materials and by employing special melting techniques in the conventional electric arc furnace. In order to obtain the optimum magnetic properties it is necessary to reduce all impurities to a very low order and after the refining operation to remove all gases and oxides since the removal of impurities is accomplished by a very severe oxidation of the metal bath. The oxides are subsequently removed by strong deoxidizing agents such as aluminum or magnesium. Careful control must be exercised at this point to eliminate all but a trace of the deoxidizer from the bath.

A second factor of primary importance in the interests of economy has been the elimination of the need for specialized tape rolling equipment. It has been demonstrated that one and two mil Deltamax can be slit to narrow tape widths as a final operation without any detrimental effect on the magnetic characteristics. Excellent rectangular hysteresis loops have been obtained on wound toroidal cores as narrow as one eighth of an inch.

In the production of wound cores, the skill of the operator is of paramount importance. The attainment of stacking factors in excess of 80 percent is made difficult since the individual convolutions of tape must be separated from each other by an insulating medium, capable of withstanding the high annealing tempera-

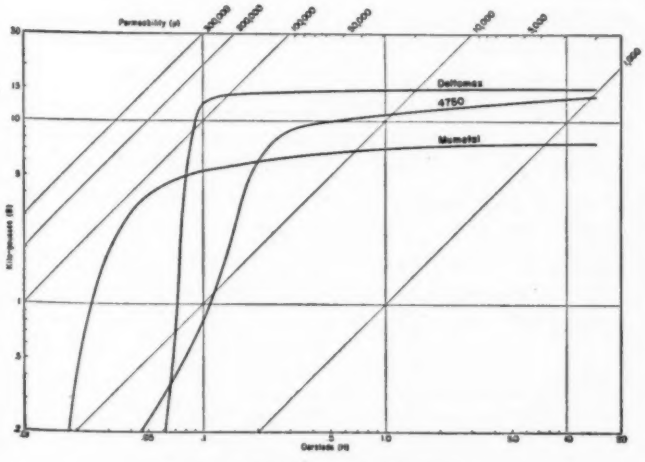


FIG. 3a. D.c. magnetization curve.

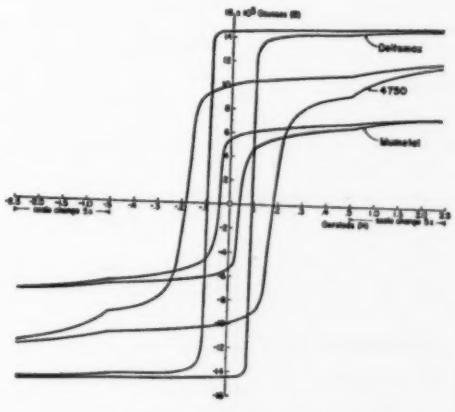


FIG. 3b. D.c. hysteresis loop.

ture, to prevent sticking. A pure grade of light magnesium oxide is most satisfactory for this purpose and can be applied dry to the tape during winding. Adherence is obtained by the light film of oil remaining on the tape from the cold rolling operation. A carefully controlled thin oxidized surface film is applied during the controlled cooling cycle and provides an effective method of reducing eddy current losses. The final anneal of the wound core is most critical and must be conducted in an atmosphere of pure dry electrolytic hydrogen. This anneal serves to relieve the severe cold rolling stresses and remove any trace of undesirable impurities remaining in the metal. In addition the high degree of preferred grain orientation necessary for the achievement of the rectangular shape of the hysteresis loop is brought about by changing the cold rolled texture into the recrystallization texture with a cube edge parallel to the direction of rolling and a cube face in the plane of the strip. The degree of purity of the metal and the amount of cold rolling are the critical factors which determine the final annealing temperature which will vary from 1000–1200°C. Excessive grain growth occurs if the critical temperature is exceeded producing a less favorable type of grain orientation with a resultant loss of the rectangular shape of the hysteresis loop. In some applications, particularly the magnetic amplifier, low coercive force is equally as important as high retentivity and a favorable compromise of the optimum values of the two properties can be obtained by controlling the total impurity content of the alloy which in turn determines the maximum annealing temperature.

A further improvement in the sharpness of the knee of the magnetization curve can be obtained by cooling from above the Curie point (500°C) in the presence of a magnetic field. Deltamax processed in this manner has attained maximum permeabilities in excess of 200,000 with a sharply defined rectangular hysteresis loop having a retentivity value of 14,800 gausses and a coercive force of 0.06 oersted, illustrated in Fig. 2.

It is interesting to compare in Fig. 3 the magnetization curves and hysteresis loops of two mil Deltamax in the form of a wound core with both Mumetal and Allegheny 4750 in similar form but processed in the conventional manner for strip products. The cores from which these curves were obtained were not annealed in a magnetic field.

The magnetic requirements of the choke cores of the mechanical rectifier have to be met under dynamic conditions and eddy currents are increased tremendously by the high speed of the reversals (one millisecond) of magnetization. Consequently, a method of a.c. testing was developed in Germany to similate the operating characteristics of the mechanical rectifier. Similar dynamic test equipment is used by the I-T-E Circuit Breaker Company in Philadelphia. Delta-*B* values (useful magnetization range) in excess of 28 kilogausses have been obtained on Deltamax cores tested on this equipment and the A.C. hysteresis loops compare favorably with the typical Permenorm 5000-Z hysteresis loops available for comparison. Typical a.c. hysteresis loops are illustrated in Fig. 4.

APPLICATION

As previously mentioned, Permenorm 5000-Z was developed for the choke cores in the mechanical rectifier and was instrumental in the successful development of this machine. It also found application in pulse transformers for radar equipment and magnetic amplifiers. Although American industry has only recently been introduced to the possibilities of this unique material, it is a foregone conclusion that many more new and useful applications will be developed as engineers become better acquainted with the characteristics of rectangular hysteresis loop materials. At the time of this writing, well over a hundred toroidal cores of Allegheny Deltamax ranging in weight from a few grams to over a pound have been distributed to over fifty industrial and research organizations to provide engineering personnel an opportunity to investigate the possibilities of this material. A better understanding of the type of small wound core particularly suited for use in magnetic amplifiers and saturable reactors can be had from an inspection of Fig. 5.

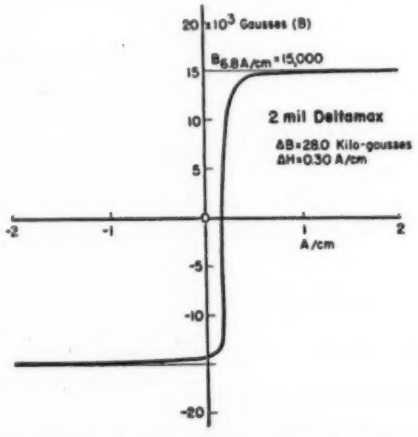


FIG. 4a. A.c. hysteresis loop (2 mil-Deltamax).

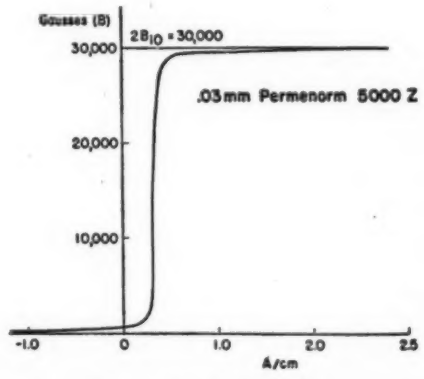


FIG. 4b. A.c. hysteresis loop (0.03 mm Permenorm 5000-Z).

It is important to mention that finished cores of this material are extremely strain sensitive and require careful handling. The use of properly designed plastic or fiber core containers is strongly recommended as a means of preserving the magnetic quality of this material.

We wish to acknowledge the many helpful suggestions contributed by Dr. E. Both and Mr. Vincent Kublin of the Squier Signal Corps Laboratory and thank Mr. W. E. Ruder of the General Electric Company for his assistance in processing the material and Mr. Otto Jensen, Manager of the Rectifier Division of the I-T-E Circuit Breaker Company, for his cooperation in performing the dynamic tests.

FIG. 5. Photograph of Deltamax cores encased in molded Nylon protective boxes with covers.

Effect of Surface Roughness on Rolling Friction

J. J. BIKERMAN
Research Laboratories, Merck and Company, Inc., Rahway, New Jersey
(Received March 24, 1949)

The minimum tilt at which bearing balls roll down an inclined plate of stainless steel is greater for rough than for smooth surfaces. The roughest surfaces (finish No. 1, average height of elevations 3×10^{-4} cm, and finish No. 2D, average height 0.9×10^{-4} cm) gave an almost quantitative agreement between this height of elevations and the height of hills calculated from a theory attributing rolling friction to surface roughness. No similar agreement was observed for fine finishes, presumably because the balls used (radii between 0.159 and 0.635 cm) were too big for the irregularities on these surfaces. The elastic deformation of the surfaces in contact was irrelevant for coarse finishes. Capillary attraction presumably did not interfere as superficial drying of the system had no effect.

Surface roughness which is the cause of Coulomb's sliding friction thus seems to cause also rolling friction at low pressures.

I. INTRODUCTION

ALTHOUGH the irregularities of the surfaces were considered long ago one of the chief causes of rolling friction[1] and an excellent surface finish was early recognized as necessary for bearing balls,[2] no quantitative study of the relation between resistance to rolling and surface roughness seems to have been published.

In the present work the minimum tilts were determined at which bearing balls rolled down inclined planes of stainless steel finished in six different fashions. The assumption that, for the conditions used, rolling friction is a result of surface roughness was confirmed by an independent roughness determination.

The old hypothesis attributing *sliding friction* (Coulomb's dry friction) to surface roughness was recently upheld in several papers.[3,4] The experiments reported here seem to prove that also *rolling friction* (under certain conditions which follow from Section IV) is caused by roughness without any "adhesion," "fusion" or "welding."

[1] O. Reynolds, Proc. Roy. Soc. London **23**, 506 (1875).
[2] Stribeck, Trans. A. S. M. E. **29**, 420 (1907).
[3] J. J. Bikerman, Rev. Mod. Phys. **16**, 53 (1944).
[4] J. J. Bikerman, Lubrication Eng. **4**, 208 (1948).

II. EXPERIMENTAL

O. Reynolds[5] determined the least tilt of a plate (of glass, wood, rubber or iron) which caused a flat iron cylinder placed upright on the plate to roll down. In the present set-up steel balls were substituted for the flat disk.

The instrument used consisted of a brass plate, 6.5 cm×3 cm, screwed to a horizontal brass shaft which could freely rotate within a brass sleeve soldered to a vertical support. Onto one end of the shaft a metal pointer could be fixed. Rotation of the shaft, i.e., tilting of the plate, caused the pointer to turn in a vertical plane in front of a circle (internal diameter 10.8 cm) divided into 360 parts and provided with a vernier. It was possible to read the tilts of the plate within 10 minutes or less but the spread of the experimental results was such that all readings were made to the nearest $\frac{1}{4}$ of a degree.

The plates used in the experiments were laid on top of the brass plate previously covered with molten wax which, on solidification, produced joints of a sufficient strength. The plates (7×5×0.08 cm) were taken from a copy of *Fabrication of U.S.S. Stainless and Heat Resist-*

[5] O. Reynolds, Phil. Trans. Royal Soc. **166**, 155 (1876).

TABLE I. Minimum tilts (in minutes of a degree) at which bearing balls rolled down steel plates. First procedure (see Section II): α_1; second procedure: α_2

Finish of the plate	Height of hills in microns	α_1 Radius of the ball in cm 0.159	0.238	0.318	0.476	0.635	α_2 Radius of the ball in cm 0.159	0.238	0.318	0.476	0.635
No. 1	3.1	217	180	154	118	97	180	128	120	86	74
No. 2D	0.8–1.0	124	100	81	61	46	87	73	62	49	48
No. 2B	0.25–0.40	117	100	84	72	56	87	64	58	33	24
No. 4	0.033–0.045	112	80	65	54	39	83	83	59	41	28
No. 6	0.084–0.10	97	75	69	46	43	72	64	51	39	22
No. 7	0.020–0.030	82	69	51	49	44	61	47	47	36	25

ing Steels (1947); they are intended to illustrate the usual finishes of 18–8 stainless steels. They are characterized as follows:

No. 1. Hot-rolled, annealed, and pickled.
No. 2D. Dull cold-rolled, i.e., cold-rolled, annealed, and pickled.
No. 2B. Bright cold-rolled, i.e., produced by final light cold-rolling.
No. 4. Standard polished.
No. 6. Tampico brushed.
No. 7. High luster polished.

A more detailed description can be found in the just mentioned publication.

Dr. S. S. Kistler, Norton Company, Worcester, Massachusetts, very kindly supplied me with roughness data on the above samples, obtained with a tracer instrument ("brush analyzer"). The root-mean-square height of their irregularities is given in the 2nd column of Table I. Two places on each plate were scanned, whence two values for the height, except the roughest plate for which both readings give identical values (3.1 microns).

In a few experiments the face of the movable spindle of a micrometer caliper was used instead of stainless steel plates.

The balls were commercial bearing balls with radii of 0.159, 0.238, 0.318, 0.476 and 0.635 cm (their diameters were $\frac{1}{8}$, $\frac{3}{16}$, $\frac{1}{4}$, $\frac{3}{8}$ and $\frac{1}{2}$ inch). There was no significant difference between the results on two sets of balls.

The balls and the plates were frequently (usually after each 10 tilt measurements) washed with first acetone and then carbon tetrachloride or first carbon tetrachloride and then acetone, and dried in air, protected from dust. The order of application of the solvents did not affect the results if about 15 min. were allowed for evaporation. In several experiments the evaporation took place in a current of air which passed first a column of anhydrous calcium sulfate and then a heated glass tubing; also this modification of the procedure had no definite effect on the minimum tilts. After washing, the balls and the plates were not touched by hand or greasy objects.

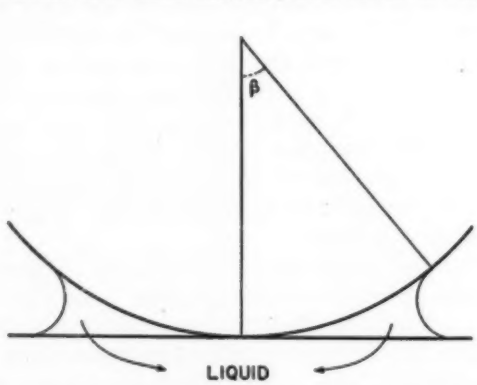


FIG. 1. Sketch for calculating the capillary attraction between a sphere and a plane if the point of contact is surrounded by liquid.

Two procedures were used for determining the minimum tilt. In the first, a ball was placed on a plate tilted at an angle x minutes to the horizon and stopped (with a piece of paper or regenerated cellulose) when it started to roll down. If the ball continued to roll as soon as the obstacle was removed, and did it 10 times in succession, the ball was removed, the tilt reduced to $(x-15)$ minutes, and so on until the ball remained immobile at least once in 10 attempts. The smallest tilt at which it could not be immobilized in this manner is denoted by α_1 in the following. In the second procedure a ball was placed onto a horizontal plate and the plate turned at a rate of about 1° per 20 seconds until the ball was set in motion. The angle read at this moment is α_2.

The results were not noticeably affected by the steadiness (or otherwise) of the experimenter's hand. In the first procedure, paper or cellulose pieces of different dimensions, which would differently transmit any vibration of the hand, yielded consistent values of α_1. In the second procedure, the tilt α_2 was the same irrespectively whether the original tilt of the plate (when the ball was being placed on it) was zero, as usual, or slightly positive or slightly negative; if, e.g., α_2 was 150 minutes, the ball could be put onto the plate tilted at -60 minutes or $+60$ minutes and still rolled down when the tilt reached about 150 minutes.

As in so many other friction experiments the reproducibility of single values of α_1 and α_2 was poor. For instance, a 0.159 cm ball gave, on the plate Finish No. 1, the following 10 successive values of α_2: 270′, 225′, 90′, 120′, 105′, 150′, 210′, 90′, 225′, 120′. The values listed in Table I are, for α_1, averages of at least 8 and, for α_2, of at least 20 determinations.

III. RESULTS

The results of the final series of experiments are summarized in Table I.

The angle α_2 for the surface of a micrometer caliper was so small that not even an approximate determination was possible using larger balls; for the smallest ball α_2 seemed to be between 15 minutes and 30 minutes.

O. Reynolds,[5] for an iron disk, 15 cm in diameter and 5 cm thick, rolling on a flat cast iron plate, found the minimum tilt to be 4 minutes ($\tan\alpha_2$ was 0.057/50).

IV. THEORY OF THE EXPERIMENT

The explanation of the above results is based in this paper mainly on the assumption that resistance to rolling is caused by surface roughness. However, before quantitative relation between the angle of rolling (α_1 or α_2) and roughness is derived, two other possible explanations are discussed. They are (a) also at the minute loads used in this work resistance to rolling is due to elastic deformation of plate and ball, and (b) separation of ball and plate is hindered by capillary attraction. The differential sliding discovered by Reynolds[5] need not be considered as Reynolds proved it to be negligible for hard materials such as iron.

1. Role of elastic deformation. H. Hertz's theory[6] gives the expression 1.23 $(P^2/E^2R)^{\frac{1}{3}}$ for the vertical displacement produced by a ball of the radius R pressed against a plate with the force P, if the Young modulus of both plate and ball is E and their common Poisson ratio is 0.3. The E of the plates used is given by the makers as 28×10^6 lb./sq. in., i.e., 1.93×10^{12} dynes/cm^2. The E of the balls is assumed to equal that of the plates (the bearing balls studied by Stribeck[2] had $E=2.08\times10^{12}$ dynes/cm^2). The greatest load P applied in this work was 8220 dynes, as the mass of the biggest balls was 8.38 g. Their R was 0.635 cm. Hence, the greatest vertical displacement was 3.72×10^{-6} cm or 0.0372μ. It is seen that the greatest indentation increased about twofold the roughness of the finishes No. 4 and No. 7 but could not have any significant effect on the finishes No. 1, 2D and 2B.

When the load on the balls is much greater than their weight, e.g., when ball bearings are used in machinery, elastic deformations greatly exceed the original surface irregularities and the resistance to rolling is not caused by the original roughness.

2. Role of capillary attraction. An approximate calculation[7] of the capillary attraction f leads to formula $f=\gamma.\ 2\pi R\sin\beta$, γ being the surface tension of the liquid (which wets the solids perfectly), R the radius of the sphere, and β the angle shown in Fig. 1. The importance of capillary attraction would be greatest for the smallest sphere. Setting $R=0.159$ cm and $\gamma=70$ dynes/cm (as of water at room temperature), f becomes $70\sin\beta$ dynes. As the mass of the smallest ball is 0.131 g, the gravitational force pulling it along the inclined plane is about $130\sin\alpha$ dynes, α being the angle of tilt. It is seen that capillary attraction may markedly influence the movement of balls as $\sin\beta$ may be as great as or greater than $\sin\alpha$.

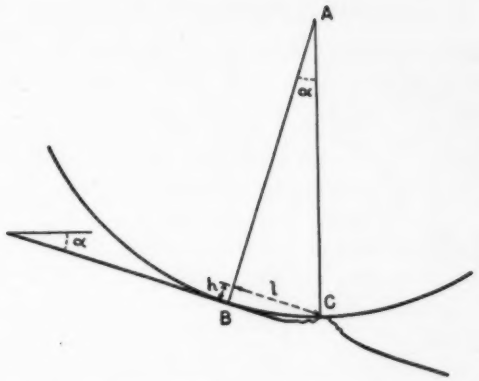


FIG. 2. Sketch for calculating the effect of surface roughness on rolling friction. α is the angle of tilt, h the height of the hill preventing the sphere from rolling down, and l the distance of the hill from AB which is normal to the inclined plate.

It is believed, however, that capillary effects were negligible because, as pointed out in Section II, the final treatment of the surface so little affected the angles of rolling. Capillary attraction presumably would be influenced by the nature of the rinsing liquid (acetone or CCl_4) and the dryness of the air in which this liquid evaporated.

3. Role of surface roughness. Surface roughness can oppose rolling down of balls because, when the plate is tilted, the ball may rest on some hills of the plate. To simplify the discussion it is assumed in the following that (a) the ball rests on one hill only and, consequently,

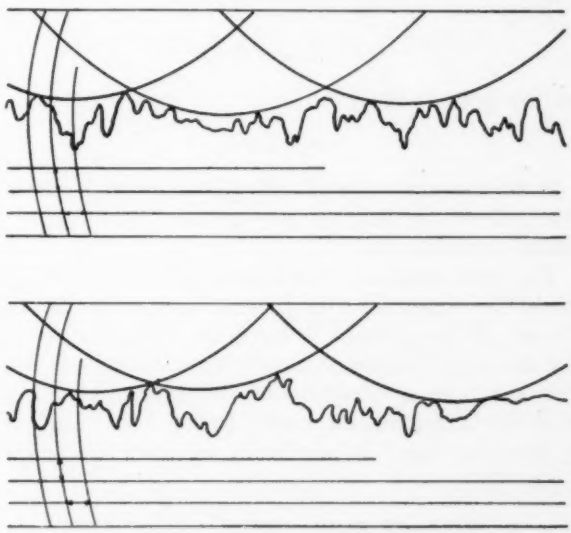


FIG. 3. Contact between smooth spheres and a plate having No. 1 finish. The irregular curves show the profile of the plate in two places. The six elliptical curves are outlines of an ½-inch ball in six positions on the plate. The distance between the parallel lines is 63.5μ in the horizontal and 12.7μ in the vertical direction.

[6] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934), p. 343.

[7] J. J. Bikerman, *Surface Chemistry for Industrial Research* (Academic Press, Inc., New York, 1947), pp. 31–32, 179.

TABLE II. Height of hills, h, and distance between hills and valleys, l, calculated from rolling friction experiments of Table I.

Finish of the plate and the height of hills in μ		$h=R(1-\cos\alpha_1)$ in μ; $l=R\sin\alpha_1$ in μ Radius of the balls in cm					$h=R(1-\cos\alpha_2)$ in μ; $l=R\sin\alpha_2$ in μ				
		0.159	0.238	0.318	0.476	0.635	0.159	0.238	0.318	0.476	0.635
No. 1 3.1	h	3.2	3.3	3.2	2.8	2.5	2.2	1.6	1.9	1.9	1.5
	l	100	125	142	163	179	83	89	111	119	137
No. 2D 0.8–1.0	h	1.0	1.0	0.89	0.76	0.57	0.51	0.55	0.51	0.48	0.63
	l	57	69	75	84	85	40	51	56	68	89
No. 2B 0.25–0.40	h	0.92	1.0	0.95	1.0	1.1	0.51	0.40	0.44	0.23	0.16
	l	54	69	78	100	103	40	44	54	46	44
No. 4 0.033–0.045	h	0.84	0.64	0.57	0.57	0.41	0.47	0.69	0.48	0.33	0.14
	l	52	55	60	75	72	38	57	55	57	52
No. 6 0.084–0.10	h	0.64	0.57	0.64	0.43	0.51	0.35	0.40	0.35	0.30	0.14
	l	45	52	64	64	79	33	44	47	54	41
No. 7 0.020–0.030	h	0.45	0.48	0.35	0.48	0.51	0.25	0.22	0.29	0.26	0.17
	l	38	48	47	68	81	28	32	43	50	46

the point of contact lies in the plane normal to the plate and including the center of the ball, and (b) that only the plate is rough while the ball is perfectly smooth. Figure 2 was drawn according to these simplifications. The ball remains stationary as long as the perpendicular through its center is situated left of the point of contact C; and when the tilt increases so much that the vertical moves to the right of C, the ball starts rolling. In Fig. 2 the vertical passes through C, and the angle α formed by it and the normal AB to the plate is the minimum tilt at which rolling starts. Figure 2 shows that the height h of the hill, on which the ball (whose radius is R) rests, is

$$h=R(1-\cos\alpha) \tag{1}$$

and its distance from AB is

$$l=R\sin\alpha. \tag{2}$$

If the average angle of rolling is α, we may conclude that hills of the height h are frequent on an area the radius of which is l.

The conventional "coefficient of rolling friction" would be equal for the present system to $(P\sin\alpha/P\cos\alpha)\cdot R$, i.e., $R\tan\alpha$, P being the weight of the ball. For the small angles measured the difference between $\sin\alpha$ and $\tan\alpha$ is smaller than the uncertainty arising from the experimental error; hence, the "coefficient of rolling friction" is practically equal to the length l of Eq. (2).

V. SURFACE ROUGHNESS, CALCULATED AND OBSERVED

Table II contains the values of h and l calculated from Eqs. (1) and (2), using α_1 for the left and α_2 for the right half of the table.

Assuming the value 20 minutes for α_2 of the micrometer face when $R=0.159$ cm, h would be 0.025μ and l, 9μ. The roughness of this surface was not measured, but the average height of the hills on a (British) micrometer anvil is about 0.015μ.[7] The cast iron plate of O. Reynolds[5] would have $h=0.1\mu$ and $l=171\mu$.

For the roughest surfaces No. 1 and No. 2D Table II reveals satisfactory agreement between the values of h calculated from minimum tilts and the root mean square heights of the hills determined with a tracer instrument.

The absence of such an agreement for the finer finishes No. 4, 6 and 7 (No. 2B is on the borderline) does not seem to be due only to the low precision of the experimental results (α_1 and α_2 having been read to the nearest $\frac{1}{4}$ of a degree) and to the elastic deformation discussed in Section IV, 1. Presumably, the roughness differences between fine finishes are not registered by the bearing balls because these balls are very big compared with the surface irregularities. It is well known that tracer needles of roughness-measuring instruments reveal more of the surface profile the thinner the needle (or the smaller the radius of curvature of its tip). A "needle" whose radius of curvature is as great as that of the balls used in all probability would be just as little sensitive to fine rugosity. On the other hand, rolling of balls with radii of 10 to 100μ would be expected to register the differences between the fine finishes also. Perhaps, the minimum tilts in the instances of No. 4, No. 6 and No. 7 are determined by the deviation of the test plates from absolute flatness rather than from absolute smoothness. The low value of α observed for the micrometer caliper supports this view. The finish of the micrometer surface appears to be

about as good as that of No. 7 but its flatness may be expected to be more perfect than that of a sample representing surface finish.

The low value of α for the micrometer caliper eliminates another conceivable explanation for the discrepancy between h and roughness values of fine finishes. The value of h includes also the height of the hills on the ball, see Section IV, 3. If these hills were considerably taller than those on No. 4, 6 and 7, the smaller elevations would not be detected in the present procedure. But, the combined height of elevations on ball and micrometer being only 0.025μ, the finish of the ball must be at least as good as that of plate No. 7.

The relation between minimum tilt and roughness, schematically shown in Fig. 2 and confirmed for the rough surfaces of Table II, is demonstrated in Fig. 3. In this figure a representative segment of each of the two profile curves recorded on plate No. 1* is shown together with outlines of the 0.159-cm ball in 3 positions on every segment. The magnification for the profile and the outline is about 50 for the horizontal and about 250 for the vertical direction; the distance between the vertical parallel curves is 63.5μ, and between the horizontal parallel lines 12.7μ.

The theory expects the vertical distance between the lowest point of the sphere and the point of its contact with a hill to be, on the average, h, i.e., in this instance 2.2–3.2μ. The horizontal distance between these two points should be equal to l, i.e., 83–100μ. Figure 3 accords with these expectations. At least for finish No. 1 surface roughness seems to account satisfactorily for the observed angles of rolling.

The systematic difference between the values of α_1 and α_2 can then be explained. At the tilt α_1 not one of 10 attempts to stop the ball was successful while at the tilt α_2 the balls rolled down at the first attempt. This means that hills of the height $R\,(1-\cos\alpha_1)$ are less frequent than those of the height $R(1-\cos\alpha_2)$; and the values of $R\sin\alpha_1$ and $R\sin\alpha_2$ in Table II agree with this conclusion.

The size of the ball has but little effect on the h of coarse plates but l increases with R in a pretty regular manner. This is a consequence of Eqs. (1) and (2); if $h=R(1-\cos\alpha)$ remains constant, $R\sin\alpha$ must increase with R as long as the problem has physical meaning. Presumably, however, the increase of l with R has not only a trigonometric explanation.

Because sliding friction still is believed by some scientists to have "molecular nature," consideration was given to the possibility of making molecular forces responsible for resistance to rolling. This possibility seems excluded, however, not only because surface roughness alone accounts for the observations but also because the cleaning of the plates adopted here proved inadequate to obtain reproducible values for the least tilt at which sliding of water drops occurred. Sliding of drops, via hysteresis of wetting, depends on molecular forces and is very sensitive to impurities on solid surfaces. Rolling of spheres does not have this sensitivity and cannot have an analogous mechanism.

ACKNOWLEDGMENT

I thank Dr. P. K. Frolich, Vice President for Research and Development, for permission to use the laboratories of Merck and Company, Inc., for this investigation and the staff of the Engineering Division for their assistance. This work could not have been completed without the much esteemed help of Dr. S. S. Kistler and Mr. R. E. Englund of Norton Company, Worcester, Massachusetts, who very kindly supplied me with surface roughness data on my samples.

* These curves were obtained through courtesy of Dr. S. S. Kistler, Norton Company.

The Demodulation of a Frequency-Modulated Carrier and Random Noise by a Discriminator*

NELSON M. BLACHMAN**
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received March 10, 1949)

With the aid of an idealized representation of the discriminator, Middleton has shown that a frequency-modulated carrier remains intelligible for smaller carrier strengths when narrow-band f-m is used and the limiter is omitted than with a limiter and/or wide-band f-m. It is, therefore, of some interest to treat the demodulation process as it *actually* occurs, in the absence of a limiter. Thus, the discriminator has been taken to consist of two selective circuits, both fed by the output of the intermediate-frequency amplifier, but peaked at different frequencies, feeding rectifiers whose outputs are subtracted. Each half of the device is treated in the manner used by Rice to determine the result of passing random noise and a sine wave through a rectifier; however, there is a correlation between the noise voltages fed to the two rectifiers. The signal output and the spectral distribution of the noise output are obtained first for quadratic rectification, then in the general case, which is then specialized to linear rectification. The results are applied to a case of rectangular i-f noise spectrum, and the signal-to-noise ratio is determined for the cases of narrow-band and wide-band f-m. These results are found to be very much like those for the idealized representation of the discriminator; all are compared, along with amplitude modulation, in Table I. The optimum signal-to-noise ratio for f-m without a limiter is found to obtain with narrow-band f-m when the discriminator is designed for no wider a band than necessary; this optimum signal-to-noise ratio differs very little from that for a-m.

I. INTRODUCTION

WITH the aid of an idealized representation of the discriminator, Middleton[1] has shown that a frequency-modulated carrier remains intelligible for smaller carrier strengths when narrow-band f-m is used and the limiter is omitted than with a limiter and/or wide-band f-m. He has obtained the spectral distribution of the output noise, as well as the signal output. The former is important in determining the output signal-to-noise ratio, since the audio system may, especially in the case of wide-band f-m, cut off the upper end of the noise spectrum, thus improving the ratio.

This idealized discriminator is a device whose output varies jointly with the instantaneous amplitude of its input and with the deviation of the instantaneous frequency of the latter from the intermediate frequency f_0. An actual discriminator generally consists of two selective circuits, both fed by the output of the i-f amplifier but peaked at different frequencies, feeding rectifiers whose outputs are subtracted. Since it cannot respond instantaneously to either frequency or amplitude, it must behave somewhat differently from the idealized discriminator. It is, therefore, of some interest to treat the demodulation process as it *actually* occurs in the absence of a limiter. We shall apply Rice's analysis of the effect of a non-linear circuit on noise plus a carrier[2] to the foregoing actual discriminator. We shall use the subscripts 1 and 2 to distinguish the outputs of the two selective circuits and their respective rectifiers. It is not necessary that the two selective circuits be distinct; they may have common components, as in the case of the Foster-Seeley discriminator.

II. QUADRATIC RECTIFICATION

Following Rice, we may commence with the relatively simple case of quadratic rectification.[3] The output voltages of the two selective circuits consists of noise and carriers of angular frequency p:

$$\left.\begin{aligned} V_1 &= P_1 \cos pt + V_{N1}, \\ V_2 &= P_2 \cos(pt+\theta_p) + V_{N2}, \end{aligned}\right\} \tag{1}$$

where $P_{1,2}$ are the strengths of the two carriers, which differ on account of the different responses of the two selective circuits, and θ_p is the difference in phase shifts through the two selective circuits at angular frequency p. These two relations (1) replace Rice's single (4.1-13).[4] If the outputs of the two rectifiers are $aV_{1,2}^2$, we have immediately the d.c. output of the discriminator from (4.5-11):

$$\tfrac{1}{2}a(P_1^2-P_2^2)+a(N_1^2-N_2^2), \tag{2}$$

where

$$N_{1,2}^2=\int_0^\infty w_{1,2}(f)df \tag{3}$$

are the mean-square noise voltages at the outputs of the two selective circuits, with $w_{1,2}(f)$ the power-spectral intensity of these noise output voltages. For a symmetric (balanced) discriminator ($w_2(f)=w_1(2f_0-f)$), $N_1=N_2$, and the second term of (2) drops out. Expression (2) represents the signal output of the discriminator, since it is the only part of the output that does not average to zero when the carrier frequency

* This article summarizes a part of the author's doctoral dissertation, "The demodulation of a frequency-modulated carrier and random noise by an f-m receiver," which appeared as Cruft Laboratory Technical Report No. 31 (March 5, 1948). This work was supported successively under OSRD Contract OEMsr-1441 and joint ONR-Signal Corps Contract N5ori-76 T. O. I, with Cruft Laboratory.
** Now at the Brookhaven National Laboratory, Upton, Long Island, New York.
[1] D. Middleton, J. App. Phys. **20**, 334 (1949). Quart. App. Math. **6**, 129 (1949). Part II is Report No. 62 (November 24, 1948).
[2] S. O. Rice, Bell Sys. Tech. J. **24**, 46 (1945), continued from **23**, 283 (1944).
[3] Reference 2, p. 125 ff.
[4] Numbers of this type will, throughout, refer to the equations in reference 2.

$f_p = p/2\pi$ has a given deviation $D = f_p - f_0$. The sinusoidal output (4.5–12) is of no interest, since, having twice the carrier frequency, it is not passed by the audio circuits.

There is, in addition, the continuous (noise) output spectrum $W_c(f)$ coming from

$$2P_1V_{N1}\cos pt - 2P_2V_{N2}\cos(pt+\theta_p) + V_{N1}^2 - V_{N2}^2,$$

which may be treated by Rice's procedure[5] if one remembers that V_{N1} and V_{N2} are derived from a common source, so that the amplitudes and phase angles of their Fourier components differ only on account of the differences in the two selective circuits. The latter introduce a phase difference $\theta(f)$ at frequency f and an amplitude ratio $(w_2/w_1)^{\frac{1}{2}}$, since $w_{1,2}(f)$ is proportional to the square of the gain of the i-f amplifier (which supplies practically all of the selectivity of the receiver) plus selective circuit. Thus, we find

$$\begin{aligned} W_c(f) &= a^2P_1^2w_1(f_p+f) + a^2P_2^2w_2(f_p+f) \\ &\quad - 2a^2P_1P_2w_1^{\frac{1}{2}}(f_p+f)w_2^{\frac{1}{2}}(f_p+f)\cdot\cos[\theta(f_p+f)-\theta_p] \\ &\quad + a^2P_1^2w_1(f_p-f) + a^2P_2^2w_2(f_p-f) \\ &\quad - 2a^2P_1P_2w_1^{\frac{1}{2}}(f_p-f)w_2^{\frac{1}{2}}(f_p-f)\cdot\cos[\theta(f_p-f)-\theta_p] \\ &\quad + a^2\int_{-\infty}^{\infty}\{w_1(f'-f)w_1(f') + w_2(f'-f)w_2(f') \\ &\quad - 2w^{\frac{1}{2}}(f'-f)w_2^{\frac{1}{2}}(f'-f)\cdot w_1^{\frac{1}{2}}(f')w_2^{\frac{1}{2}}(f')\cos[\theta(f'-f) \\ &\qquad\qquad - \theta(f')]\}df', \end{aligned} \tag{4}$$

where $w_{1,2}(-f)$ and $\theta(-f)$ are defined by

$$w_{1,2}(-f) = w_{1,2}(f), \quad \theta(-f) = -\theta(f).$$

The total mean-square discriminator noise output is $\int_0^\infty W_c(f)df$. Since $w_{1,2}(f)$ will be different from zero only in the vicinity of the intermediate frequency f_0, we may equally well take $w_{1,2}(f)$ to be zero for negative f and insert a factor 2 in front of the integral in (4). In the case of a symmetric discriminator, one for which $w_1(f_0+f) = w_2(f_0-f)$, it can be seen that the first two terms in the integral in (4) contribute equal amounts to $W_c(f)$.

Equation (4) gives the continuous spectrum of the discriminator output when the carrier is held at a constant frequency f_p. We may consider the situation in which f_p varies slowly and sinusoidally about the intermediate frequency f_0, i.e., the carrier has deviation $D = f_p - f_0$ with D varying sinusoidally at a low frequency. If the discriminator output is analyzed with the aid of a set of similar, properly terminated band pass filters passing adjacent ranges of frequencies, the power dissipated in the load of a filter of pass band width df situated at frequency f is, when the carrier deviation is D, (4) times df, which we may write as

[5] Reference 2, p. 129.

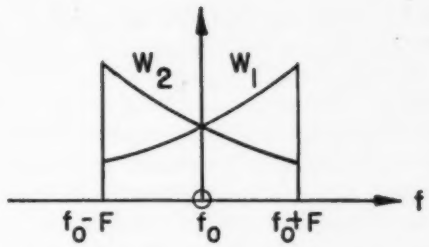


FIG 1. The assumed power spectra of the noises fed to the two discriminator rectifiers.

$W_c(f,D)df$. If, now, D varies slowly and sinusoidally with amplitude D_0, the fraction of the time it is D within dD is $(D_0^2 - D^2)^{-\frac{1}{2}}dD/\pi$ for $|D| < D_0$ and 0 for $|D| > D_0$. Integrating over all values of D, we obtain the discriminator output noise spectrum

$$W(f) = \pi^{-1}\int_{-D_0}^{D_0}(D_0^2 - D^2)^{-\frac{1}{2}}W_c(f,D)dD. \tag{5}$$

The same method is, of course, applicable when D varies slowly and not sinusoidally; if the probability that the deviation is D within dD is $P(D)dD$, the discriminator output noise spectrum is $W(f) = \int_{-\infty}^{\infty}P(D) \times W_c(f,D)dD$. If D does not vary slowly, the discriminator output spectrum will have a very toothy shape which depends strongly on the modulating frequency. However, (5) is a smoothed version of the exact result.

Application

We shall now apply the foregoing results for quadratic rectification to a very simple example. We shall suppose that $\theta(f)$ is a constant, at least over the pass band of the i-f amplifier, so that it drops out of (4), and that

$$\left.\begin{aligned} w_1(f_0+f) &= \tfrac{1}{2}F^{-1}R^{-1}N^2(1+qf)^2 \quad \text{for } -F<f<F, \\ &= 0 \quad \text{for } -f_0<f<-F,\ F<f; \\ w_2(f_0+f) &= \tfrac{1}{2}F^{-1}R^{-1}N^2(1-qf)^2 \quad \text{for } -F<f<F, \\ &= 0 \quad \text{for } -f_0<f<-F,\ F<f; \\ \theta(f) &= \text{constant}; \end{aligned}\right\} \tag{6}$$

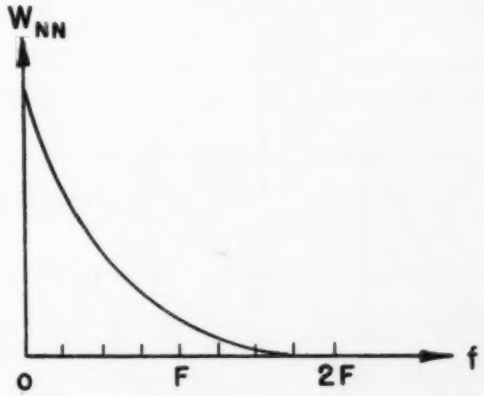


FIG. 2. The part of the discriminator output noise spectrum arising from beats between the input noise and itself.

with $R=1+\frac{1}{3}q^2F^2$, so that (3) is satisfied with $N_1=N_2=N$ the r.m.s. noise voltage fed to either rectifier; $2F$ is the i-f band width, and q is a (positive) constant of the discriminator such that $qF<1$. Thus, $1<R<4/3$; R cannot depart greatly from unity. These w_i are sketched in Fig. 1. They correspond to a discriminator characteristic, which is proportional to $w_1^{\frac{1}{2}}-w_2^{\frac{1}{2}}$, that is linear over the range of the i-f pass band.

It is convenient to separate the spectrum (4) into two parts:

$$W_e(f)=W_{CN}(f)+W_{NN}(f).$$

The former consists of the first six terms of (4), which are due to beats between the carrier and the noise; the latter is the integral in (4), which arises from beats between the noise voltage and itself. Substituting (6) into the integral of (4), we have, on integrating,

$$\begin{aligned} W_{NN}(f)&=\tfrac{2}{3}F^{-2}R^{-2}q^2a^2N^4(2F-f)^3 \quad \text{for } 0<f\leqslant 2F, \\ &=0 \quad \text{for } 2F\leqslant f, \end{aligned} \tag{7}$$

which plots as the cubic parabola shown in Fig. 2. W_{NN} is the portion of W_e that predominates when the carrier is weak.

Integration of (7) gives the mean-square total noise due to noise-noise beats:

$$\int_0^\infty W_{NN}(f)df=\frac{8}{3}R^{-2}q^2a^2N^4F^2. \tag{8}$$

If the audio system is a low pass filter cutting off at f_a, and f_a is much less than the i-f bandwidth $2F$, as is the case with wide band f-m, the mean-square audible output noise due to noise-noise beats is

$$\int_0^{f_a<<2F} W_{NN}(f)df=\frac{16}{3}R^{-2}q^2a^2N^4Ff_a. \tag{9}$$

$W_{CN}(f)$ depends on the carrier deviation D, which, in turn, determines the relative values of P_1 and P_2. Because the same selective circuits determine these two carrier amplitudes as determine the noise spectra (6), we can write

$$\left.\begin{aligned} P_1&=(1+qD)P, \\ P_2&=(1-qD)P, \end{aligned}\right\} \tag{10}$$

with P a constant which may be considered as the average carrier amplitude. Substituting (10) into the first six terms of (4), we get

$$W_{CN}(f)=2F^{-1}R^{-1}P^2q^2a^2N^2[(2D+f)^2+(2D-f)^2], \tag{11}$$

from which the first term in brackets is to be omitted whenever $f>F-D$ and the second whenever $f>F+D$. (Thus, $W_{CN}(f)=0$ for $f>F+D$.) W_{CN} is shown for several values of D in Fig. 3. The dashed lines represent the two terms, respectively, corresponding to the first three and second three terms on the right side of (4).

Substituting (11) into (5) and integrating, we obtain for the case of slow sinusoidal frequency modulation of maximum deviation D_0

$$\left.\begin{aligned} W_{CN}(f)&=4F^{-1}R^{-1}P^2q^2a^2N^2(f^2+2D_0^2) \\ &\qquad \text{for } 0\leqslant f\leqslant F-D_0, \\ &=4F^{-1}R^{-1}P^2q^2a^2N^2\{(f^2+2D_0^2) \\ &\quad \times\arccos[(f-F)/D_0]-2(2f-F)[D_0^2 \\ &\quad -(f-F)^2]^{\frac{1}{2}}\} \text{ for } F-D_0\leqslant f\leqslant F+D_0, \\ &=0 \qquad \text{for } F+D_0\leqslant f. \end{aligned}\right\} \tag{12}$$

Figure 4 shows Eq. (12). Notice that it is a synthesis of spectra like those shown in Fig. 3, each weighted according to (5).

W_{CN} is the part of the output noise spectrum W_e which predominates when the carrier is strong. It is seen that the spectral distribution of the noise output improves with increasing carrier strength, for it changes from something like Fig. 2 when the carrier is weak, in which only a relatively small part of the spectrum will be cut off by the audio system (low pass filter), to something like Fig. 4 when the carrier is strong, where relatively more is cut off by the audio system in the case of wide band f-m. The total mean-square noise due

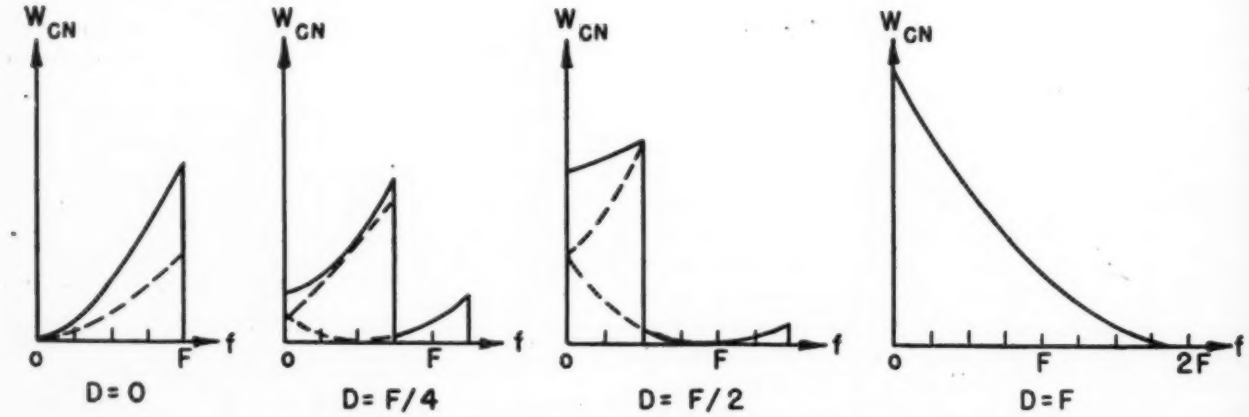


FIG. 3. The part of the discriminator output noise spectrum due to beats between the input noise and carrier shown for several values of the deviation, assumed constant.

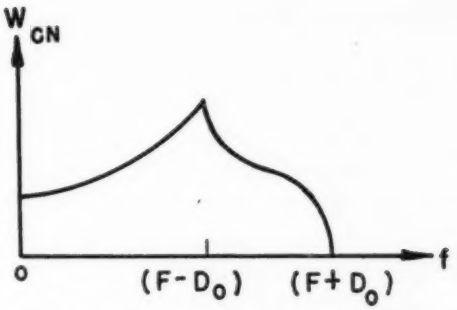


FIG. 4. $W_{CN}(f)$ with slow sinusoidal modulation.

to carrier-noise beats is

$$\int_0^\infty W_{CN}(f)df=\frac{4}{3}R^{-1}P^2q^2a^2N^2(F^2+\tfrac{3}{2}D_0^2). \quad (13)$$

If the cut-off frequency of the audio system f_a is much smaller than $F-D_0$, as in the case of wide band f-m, the total mean-square audible noise arising from carrier-noise beats becomes

$$\int_0^{f_a<<F-D_0} W_{CN}(f)df=8F^{-1}R^{-1}P^2q^2a^2N^2D_0^2f_a. \quad (14)$$

Substituting (10) into (2), we obtain the signal output voltage of the discriminator

$$2P^2qaD, \quad (15)$$

from which it is clear that the discriminator output contains a component proportional to the deviation of the carrier. Thus, the present discriminator, using (6) and quadratic rectification, demodulates the signal without distortion and without any effect due to the noise. If D varies sinusoidally with maximum value D_0, its mean-square value is $\frac{1}{2}D_0^2$, and the mean-square signal output voltage is

$$2P^4q^2a^2D_0^2. \quad (16)$$

The important quantity in the case of narrow band f-m is the mean-square signal-to-total-noise ratio, the ratio of (16) to the sum of (8) and (13),

$$6R^2(D_0/F)^2x^2/[2+(2+3D_0^2/F^2)Rx],$$

in which $x=P^2/2N^2$ is the mean-square carrier-to-noise ratio. If $D_0=3^{-\frac{1}{2}}F$, which is about the largest value D_0 may have if the f-m carrier is not to be distorted by the cutting off of important side-bands by the i-f amplifier, this mean-square signal-to-total-noise ratio becomes

$$2R^2x^2/(2+3Rx). \quad (17)$$

One should remember that this ratio relates to the mean-square output noise *during* modulation of the carrier; it is not correct to compare it with the ratio of mean-square signal during modulation to mean-square noise without modulation, which would be given by the ratio of (16) to the sum of (8) and (13) with D_0 taken to be 0 in the latter. Measurement of the ratio (17) requires the determination of the mean-square signal output with the aid of a sharp filter that passes only the signal, plus the measurement of the mean-square total (signal and noise) output, from which the mean-square noise output is obtained by subtraction of the mean-square signal output.

The ratio (17) can be maximized with respect to discriminator design by maximizing R, i.e., giving it a value closer to 4/3 than 1. This is done by maximizing qF, i.e., by designing the discriminator to cover the i-f pass band and no more.

The mean-square i-f noise N^2 is proportional to the i-f band width $2F$; hence, the carrier-to-noise ratio x *varies inversely with band width.* From this the well known conclusion follows that, for a maximum signal-to-noise ratio (17), the i-f bandwidth $2F$ should be as small as possible, *viz.*, no larger than twice the highest audio frequency to be transmitted, presumably f_a.

If the upper audio cut-off f_a is much less than $F-D_0$, the mean-square output signal-to-noise ratio is the ratio of (16) to the sum of (9) and (14), *viz.*,

$$\tfrac{3}{2}(F/f_a)(D_0/F)^2R^2x^2/[1+3(D_0/F)^2Rx],$$

which, for $D_0=3^{-\frac{1}{2}}F$, becomes

$$\tfrac{1}{2}(F/f_a)R^2x^2/(1+Rx). \quad (18)$$

Since (18) is similar in form to (17), it leads to the same conclusions concerning discriminator design and band width, too, for the factor F is not sufficient to alter them.

We may compare these mean-square signal-to-noise ratios with that for quadratic detection of an amplitude-modulated carrier with a rectangular i-f noise spectrum,[6]

$$2m^2x^2/[1+(2+m^2)x],$$

where m is the modulation index, which may be set equal to 100 percent, giving us

$$2x^2/(1+3x). \quad (19)$$

This ratio is of the same form as (17) varying as x^2 for small x and as x for large x. It is 0.5 db better with x small and 1.3 db worse with x large than (17) for $R=4/3$.

If the detector is followed by an audio low pass filter cutting off at $f_a\ll 2F$, the mean-square signal-to-noise ratio is[6]

$$2(F/f_a)m^2x^2/[1+(2+m^2)x],$$

which becomes

$$2(F/f_a)x^2/(1+3x) \quad (20)$$

for $m=100$ percent. Expression (20) is to be compared

[6] R. E. Burgess, "The rectification of signal and noise by linear and square-law detectors," Radio Research Board Report C93, England (March 24, 1944). J. R. Ragazzini, Proc. I. R. E. 30, 277 (1942). D. Middleton, "Some general results in the theory of noise through non-linear devices," Quart. App. Math. 5, 445 (1948); "Rectification of a sinusoidally modulated carrier in the presence of noise," Proc. I. R. E. 36, 1467 (1948).

with (18), to which it behaves similarly; it is 3.5 db better for small x and 0 db better for large x than (18) for $R=4/3$. Except for the factor (F/f_a), (18) and (20) behave just like (17) and (19).

III. GENERAL RECTIFICATION

In the case of the general rectifier (non-linear device), Rice's characteristic-function method[7] is readily applied to the problem of finding the discriminator signal and noise output spectrum. Here, if $I_1(t)$ and $I_2(t)$ are the outputs of the two rectifiers, the correlation function of the discriminator output

$$\Psi(\tau)=\langle[I_1(t)-I_2(t)][I_1(t+\tau)-I_2(t+\tau)]\rangle_{Av}$$
$$=\Psi_{11}+\Psi_{22}-\Psi_{12}-\Psi_{21}, \quad (21)$$

where, for example,

$$\Psi_{12}(\tau)=\langle[I_1(t)I_2(t+\tau)]\rangle_{Av}.$$

Ψ_{11} and Ψ_{22} are exactly like the single correlation function arising in the a-m case treated by Rice, and Ψ_{12} and Ψ_{21} are similar.

The inputs to the rectifiers are still given by Eqs. (1). Because of the amplitude difference and relative phase shift in the two selective circuits, the characteristic function for the signal portion of the discriminator input corresponding to Ψ_{12}, for example, is

$$g_{s12}(u,v,\tau)=J_0([P_1^2u^2+P_2^2v^2+2P_1P_2uv\cos(p\tau+\theta_p)]^{\frac{1}{2}}),$$

rather than (4.9-2), while that for the noise portion of the input is

$$g_{N12}(u,v,\tau)=\exp[-\tfrac{1}{2}N_1^2u^2-\tfrac{1}{2}N_2^2v^2-\psi_{12}(\tau)uv],$$

rather than (4.8-3). Rice's ψ_0 is the mean-square noise input voltage, the same as $N_{1,2}^2$; $\psi_{12}(\tau)$ is the cross-correlation function for the two noise input voltages:

$$\psi_{12}(\tau)=\langle V_{N1}(t)V_{N2}(t+\tau)\rangle_{Av},$$

which may be expressed as

$$\psi_{12}(\tau)=\int_0^\infty w_1^{\frac{1}{2}}(f)w_2^{\frac{1}{2}}(f)\cos[2\pi f\tau+\theta(f)]df, \quad (22)$$

rather than (2.1-6). Thus,

$$\Psi_{12}(\tau)=\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}(k!)^{-1}\psi_{12}{}^k(\tau)h_{nk}{}^{(1)}h_{nk}{}^{(2)}\epsilon_n$$
$$\times\cos n(p\tau+\theta_p), \quad (23)$$

where $\epsilon_0=1$ and $\epsilon_n=2$ for $n\geqslant 1$, $h_{nk}{}^{(1,2)}$ being Rice's h_{nk} computed for the two halves of the discriminator, respectively;

$$h_{nk}{}^{(1,2)}=\tfrac{1}{2}(n!)^{-1}(N_{1,2}{}^2/2)^{(\nu-k)/2}x_{1,2}{}^{n/2}$$
$$\times[\Gamma(\nu+1)/\Gamma([2-k-n+\nu]/2)]$$
$$\cdot{}_1F_1([k+n-\nu]/2;n+1;-x_{1,2}) \quad (24)$$

for νth-law rectification, with $x_{1,2}=P_{1,2}{}^2/2N_{1,2}{}^2$.

Each of the terms of the double summation (23) can be given a physical interpretation. The terms for $k=0$ represent the discrete output components; that for $n=0$ represents the d.c. component of the output, while that for $n=n$ represents an output component of frequency nf_p, i.e., n times the carrier frequency. All other terms correspond to the continuous (noise) output spectrum; the n,k term represents the $(n+k)$th-order beats between the noise (entering k times) and the carrier (entering n times).

We obtain the other Ψ_{ij} from (23) by obvious changes of subscripts and indices and, hence, by (21), Ψ. It follows that the d.c. discriminator output is

$$(h_{00}{}^{(1)2}+h_{00}{}^{(2)2}-h_{00}{}^{(1)}h_{00}{}^{(2)}$$
$$-h_{00}{}^{(2)}h_{00}{}^{(1)})^{\frac{1}{2}}=h_{00}{}^{(1)}-h_{00}{}^{(2)}. \quad (25)$$

In fact, the d.c. outputs of the two rectifiers separately are $h_{00}{}^{(1)}$ and $h_{00}{}^{(2)}$.

To obtain the correlation function $\Psi_c(\tau)$ of the continuous part of the output spectrum, we omit the terms for $k=0$; outside of the term giving the d.c. output, these terms are of no interest on account of their high frequencies. By the Wiener-Khintchine theorem, the continuous (noise) output spectrum is

$$W_c(f)=4\int_0^\infty\Psi_c(\tau)\cos 2\pi f\tau d\tau$$
$$=4\int_0^\infty(\Psi_{c11}+\Psi_{c22}-\Psi_{c12}-\Psi_{c21})\cos 2\pi f\tau d\tau$$
$$=(\text{say})W_{c11}+W_{c22}-W_{c12}-W_{c21}. \quad (26)$$

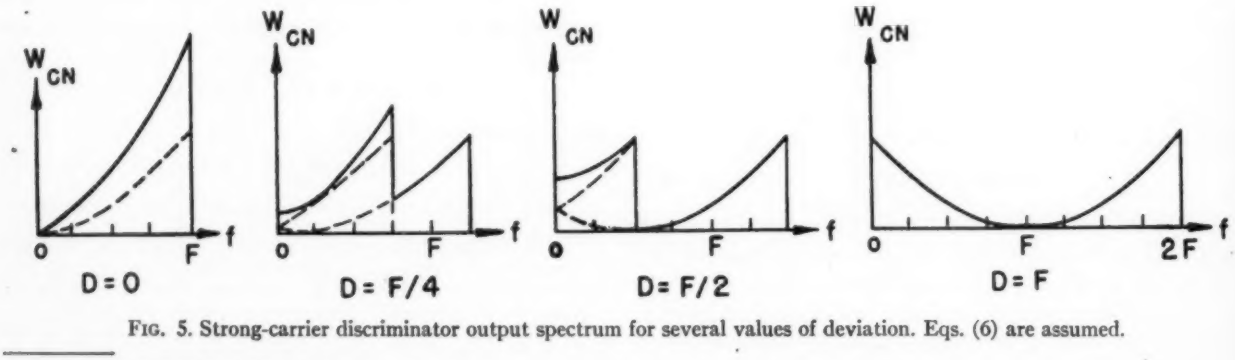


FIG. 5. Strong-carrier discriminator output spectrum for several values of deviation. Eqs. (6) are assumed.

[7] Reference 2, Sections 4.8, 4.9.

Thus, for example,

$$W_{c12}(f)=\sum_{n=0}^{\infty}\sum_{k=1}^{\infty}2\epsilon_n(k!)^{-1}h_{nk}^{(1)}h_{nk}^{(2)}\times[G_{nk12}(nf_p-f)+G_{nk12}(nf_p+f)], \quad (27)$$

where

$$G_{nk12}(f)=\int_0^{\infty}\psi_{12}{}^k(\tau)\cos(2\pi f\tau+n\theta_p)d\tau. \quad (28)$$

(Cf. (4.9–12,13).) The other W_{cij} and G_{nkij} are readily obtained by appropriate changes of subscripts, etc.

As a first approximation to $W_{c12}(f)$ for audio frequencies, we may take the sum of the two terms $n=1=k$ and $n=0$, $k=2$. From (22) and the Fourier integral theorem, it follows that

$$G_{11\ 12}(f)=\tfrac{1}{4}w_1{}^{\frac{1}{2}}(f)w_2{}^{\frac{1}{2}}(f)\cos[\theta(f)-\theta_p].$$

Similarly,

$$G_{02\ 12}(f)=\frac{1}{8}\int_{-\infty}^{\infty}w_1{}^{\frac{1}{2}}(f'-f)w_2{}^{\frac{1}{2}}(f'-f)w_1{}^{\frac{1}{2}}(f')w_2{}^{\frac{1}{2}}(f')\times\cos[\theta(f'-f)-\theta(f')]df'.$$

Substituting these G's into (27), we have our approximation for W_{c12}. Obtaining the other W_{cij} by appropriate alterations, we have from (26) finally

$$\begin{aligned}W_c(f)&=h_{11}^{(1)2}w_1(f_p+f)+h_{11}^{(2)2}w_2(f_p+f)\\&\quad-2h_{11}^{(1)}h_{22}^{(2)}w_1{}^{\frac{1}{2}}(f_p+f)w_2{}^{\frac{1}{2}}(f_p+f)\cos[\theta(f_p+f)-\theta_p]\\&\quad+h_{11}^{(1)2}w_1(f_p-f)+h_{11}^{(2)2}w_2(f_p-f)\\&\quad-2h_{11}^{(1)}h_{22}^{(2)}w_1{}^{\frac{1}{2}}(f_p-f)w_2{}^{\frac{1}{2}}(f_p-f)\cos[\theta(f_p-f)-\theta_p]\\&\quad+\frac{1}{8}\int_{-\infty}^{\infty}\{h_{02}^{(1)2}w_1(f'-f)w_1(f')+h_{02}^{(2)2}w_2(f'-f)w_2(f')\\&\quad-2h_{02}^{(1)}h_{02}^{(2)}w_1{}^{\frac{1}{2}}(f'-f)w_2{}^{\frac{1}{2}}(f'-f)w_1{}^{\frac{1}{2}}(f')w_2{}^{\frac{1}{2}}(f')\\&\quad\times\cos[\theta(f'-f)-\theta(f')]\}df'. \quad (29)\end{aligned}$$

Note that (25) and (29) check with (2) and (4) for the case of quadratic rectification ($\nu=2$), for which (29) is exact.

IV. LINEAR RECTIFICATION

For linear rectification, $\nu=1$, and (24) gives us

$$\left.\begin{aligned}h_{00}&=(2\pi)^{-\frac{1}{2}}Ne^{-x/2}[(1+x)I_0(x/2)+xI_1(x/2)],\\h_{11}&=\tfrac{1}{2}(x/\pi)^{\frac{1}{2}}e^{-x/2}[I_0(x/2)+I_1(x/2)],\\h_{02}&=(2\pi)^{-\frac{1}{2}}N^{-1}e^{-x/2}I_0(x/2),\end{aligned}\right\} \quad (30)$$

in which $x=P^2/2N^2$ and I_0 and I_1 are Bessel functions of "imaginary argument." For small carrier-to-noise ratio ($x\ll1$), Eqs. (30) become approximately

$$\left.\begin{aligned}h_{00}&=(2\pi)^{-\frac{1}{2}}N(1+x/2)\\&=(2\pi)^{-\frac{1}{2}}N+\tfrac{1}{4}(2\pi)^{-\frac{1}{2}}P^2/N,\\h_{11}&=\tfrac{1}{2}(x/\pi)^{\frac{1}{2}},\\h_{02}&=(2\pi)^{-\frac{1}{2}}N^{-\frac{1}{2}},\end{aligned}\right\} \quad (31)$$

TABLE I. Comparison of mean-square signal-to-noise ratios for f-m without a limiter and a-m.

System	Demodulator	Weak carrier ($x\ll1$): Narrow band ($f_a\gg F$)	Weak carrier ($x\ll1$): Wide band ($f_a\ll F$)	Strong carrier ($x\gg1$): Narrow band ($f_a\gg F$)	Strong carrier ($x\gg1$): Wide band ($f_a\ll F$)
f-m	Quadratic discriminator	R^2x^2	$\frac{1}{2}(F/f_a)R^2x^2$	$\frac{2}{3}Rx$	$\frac{1}{2}(F/f_a)Rx$
f-m	Linear discriminator	$2R^2x^2$	$(F/f_a)R^2x^2$	Rx	$2(F/f_a)Rx$
f-m	Idealized discriminator	$\frac{1}{4}\pi x^2$	—	x	$2(F/f_a)x$
a-m	Quadratic detector	$2x^2$	$2(F/f_a)x^2$	$\frac{2}{3}x$	$\frac{2}{3}(F/f_a)x$
a-m	Linear detector	$1.83x^2$	$1.83(F/f_a)x^2$	x	$(F/f_a)x$

$2F$ = i-f band width, f_a = audio upper cut-off frequency, x = carrier-to-noise ratio $\propto 1/F$, $D_0=3^{-\frac{1}{2}}F$ for f-m, $m=100$ percent for a-m, $1<R<4/3$.

while for a large carrier-to-noise ratio ($x\gg1$) we have approximately

$$h_{00}=P/\pi, \quad h_{11}=1/\pi, \quad h_{02}=1/\pi P. \quad (32)$$

Application

We may apply the foregoing results to the case of (6) (Fig. 1), treating, in particular, the two cases of very small and very large carrier-to-noise ratio. In the former case, the continuous-spectral output of the discriminator is given by (29) with (31). Dividing (29) into noise-noise beats, $W_{NN}(f)$, and carrier-noise beats, $W_{CN}(f)$, we see that the former is represented by the first six terms and the latter by the integral. Here the significant portion of the output noise spectrum is

$$\begin{aligned}W_c(f)=W_{NN}(f)&=(q^2N^2/24\pi F^2R^2)(2F-f)^3 \quad \text{for } 0<f\leqslant 2F \quad (33)\\&=0 \quad \text{for } 2F\leqslant f.\end{aligned}$$

It has the same shape as (7) shown in Fig. 2 because H_{02} is independent of P for small x. Similarly, because h_{11} varies as P for small x, the carrier-noise spectrum $W_{CN}(f)$, though negligible, has the same shape as (10) or (11), shown in Fig. 3 or 4.

The total mean-square output noise is, hence,

$$\int_0^{\infty}W_c(f)df=q^2N^2F^2/6\pi R^2. \quad (34)$$

(Cf. (8).) If the audio system does not pass frequencies above $f_a\ll2F$, as is the case with wide band f-m, the mean-square audible output noise is

$$\int_0^{f_a<<2F}W_c(f)df=q^2N^2Ff_a/3\pi R^2. \quad (35)$$

Substituting (10) into (25) and using (31), we obtain the discriminator signal output voltage,

$$(2\pi)^{-\frac{1}{2}}P^2N^{-1}qD.$$

(Cf. (15).) It follows as before that the demodulation

is distortionless, and that, if the carrier is frequency modulated sinusoidally with maximum deviation D_0, the mean-square signal output is

$$P^4q^2D_0^2/4\pi N^2. \tag{36}$$

(Cf. (16).) Thus, the mean-square signal-to-noise ratio for narrow band f-m is the ratio of (36) to (34), *viz.*,

$$6R^2(D_0/F)^2x^2$$

or, with $D_0=3^{-\frac{1}{2}}F$, which is about the largest value D_0 may have if the signal is not to be distorted in the i-f amplifier,

$$2R^2x^2. \tag{37}$$

It should be kept in mind that R is a parameter of the discrimator whose value lies between 1 and 4/3; x is the ratio of carrier power to noise power at the i-f amplifier output, so to speak, if the i-f amplifier is taken to include the selective circuits of the discriminator.

For the case of wide band f-m the mean-square signal-to-noise ratio is the ratio of (36) to (35),

$$3(F/f_a)(D_0/F)^2R^2x^2,$$

which, with $D_0=3^{-\frac{1}{2}}F$, reduces to

$$(F/f_a)R^2x^2. \tag{38}$$

Turning now to the case of large carrier-to-noise ratio x, we use (32) rather than (31). Here the carrier-noise beats far outweigh the noise-noise, which we neglect. Substituting (6) into (29) and using (32), we find

$$\int_0^\infty W_c(f)=W_{cN}(f)=(2q^2N^2/\pi^2FR)[(D+f)^2 +(D-f)^2], \tag{39}$$

from which the first term in brackets is to be omitted whenever $f>F-D$ and the second, also, whenever $f>F+D$. Equation (32) is shown for several values of D in Fig. 5. The dashed lines represent the two terms of (39), respectively.

Substituting (39) into (5) and integrating, we obtain for the case of slow sinusoidal modulation with maximum frequency deviation D_0

$$\left.\begin{aligned} W_c(f)&=(2q^2N^2/\pi^2FR)(2f^2+D_0^2) \\ &\qquad\qquad \text{for } 0<f\leqslant F-D_0, \\ &=(2q^2N^2/\pi^2FR)\{(2f^2+D_0^2)\arccos[(f-F) \\ &\qquad /D_0]-(3f+F)[D_0^2-(f-F)^2]^{\frac{1}{2}}\} \\ &\qquad\qquad \text{for } F-D_0\leqslant f\leqslant F+D_0, \\ &=0 \qquad\qquad \text{for } F+D_0\leqslant f, \end{aligned}\right\} \tag{40}$$

a plot of which resembles Fig. 4. The total mean-square noise represented by (40) is

$$\int_0^\infty W_c(f)df=4q^2N^2F^2/3\pi^2R; \tag{41}$$

the amount passed by a low pass filter cutting off at $f_a\ll F-D_0$ is

$$\int_0^{f_a<<F-D_0} W_c(f)df=2D_0^2q^2N^2f_a/\pi^2FR. \tag{42}$$

The signal output voltage is now $(P_1-P_2)/\pi$, which, by (10), is

$$(2/\pi)qPD,$$

whence the mean-square signal output for sinusoidal frequency modulation with maximum deviation D_0 is

$$(2/\pi^2)q^2P^2D_0^2. \tag{43}$$

Thus, the mean-square signal-to-noise ratio applicable to narrow band f-m, is the ratio of (43) to (41),

$$3(D_0/F)^2Rx$$

or, with $D_0=3^{-\frac{1}{2}}F$,

$$Rx. \tag{44}$$

The mean-square signal-to-noise ratio applicable to wide band f-m is the ratio of (43) to (42),

$$2(F/f_a)Rx, \tag{45}$$

independent of D_0.

When the carrier-to-noise ratio x is of the order of unity, both the carrier-noise and the noise-noise spectra are significant. The latter has the shape of Fig. 2, the former that of Figs. 3 and 5, or 4. Here it is necessary in evaluating (29) to use the exact expressions for the h_{nk}, (30). Although in the extreme cases considered above the discriminator demodulates distortionlessly, in the general case there is distortion. Otherwise, the general behavior of the discriminator with linear rectification is very much like that with quadratic.

In Table I all of our signal-to-noise-ratio results are summarized for the two cases of very small and very large carrier-to-noise ratio x. The corresponding signal-to-noise ratios are shown, also, for quadratic and linear detection of a 100 percent-amplitude-modulated carrier with a rectangular i-f frequency-response characteristic of width $2F$,[6] as well as for the idealized discriminator (Section I) without a limiter.[8]

These signal-to-noise ratios are exhibited for the two cases where the i-f band width $2F$ is less than, and much greater than the frequency range f_a of the audio system, which, for f-m, represent, respectively, the narrow band and wide band cases. With a-m, of course, it is undesirable to use a wide band ($F>f_a$); although it may seem desirable because of the factor (F/f_a), one must remember that the mean-square i-f noise is proportional to F, and so the i-f carrier-to-noise ratio x varies inversely with F, while f_a is, presumably, fixed, being the highest frequency signal component to be transmitted. This conclusion applies likewise to f-m; without a limiter

[8] N. M. Blachman, J. App. Phys. **20**, 38 (1949); also Cruft Laboratory Technical Report No. 31. See also, references 1.

it is undesirable from the point of view of the signal-to-noise ratio to use wide band f-m.

Middleton has shown[1,9] that, if a limiter is used and provided that the carrier-to-noise ratio is very high, wide band f-m is able to give a considerably better signal-to-noise ratio than any of the other systems considered here. However, in order for this system to be advantageous, it is necessary that the signal be already quite intelligible; otherwise the use of a limiter worsens the signal-to-noise ratio. Unfortunately, it has not been practically possible to analyze the operation of an actual discriminator, such as those with which we have dealt above, preceded by a limiter, because of the difficulty of treating cascaded non-linear circuits connected by a selective circuit.

From Table I we see that the signal-to-noise ratios for all three discriminators, as well as both detectors, differ very little; for small carrier-to-noise ratio x, all vary as x^2, while for large x, all vary as x. The spectral distribution of the output noise in the case of fm will be roughly a sum of the noise-noise spectrum, shown in Fig. 2, and the carrier-noise spectrum, shown in Fig. 4, with the former predominating for small x and the latter for large. The effect of the factors R^2 and R appearing in Table I is at most to improve the signal-to-noise ratio by 2.4 db for small x and 1.2 db for large x, if the discriminator is designed so that its characteristic is linear over only the pass band of the i-f amplifier.

It is a pleasure to acknowledge the helpful guidance of Professor J. H. Van Vleck, Dr. David Middleton, Professor P. E. Le Corbeiller, and Professor E. L. Chaffee.

[9] See also S. O. Rice, Bell Sys. Tech. J. **27**, 109 (1948).

Effective Use of Collimating Apertures in Small-Angle X-Ray Diffraction Cameras*

ORVIL E. A. BOLDUAN** AND RICHARD S. BEAR
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received March 23, 1949)

The use of simple, aperture-collimated cameras in x-ray diffraction studies at small angles is of value in investigating the molecularly large structural features of colloidal systems, particularly when pinhole collimation is required. To make effective use of available x-radiation one must (1) consider the initial limitations imposed on a camera by the purpose for which it is to be employed, and (2) choose remaining camera dimensions so that the registering film may receive as much radiation as possible.

The initial conditions are three in number, involving the desired pattern magnification, the necessary angular or large-spacing resolution, and the requisite guarding against beam spill-over, which together essentially limit the specimen-to-film distance, the beam diameter at the film, and the diffraction area permitted to be unguarded but covered by a beam stop. Other camera dimensions are chosen so that the film center is as close to, while viewing as much of, the radiation source as possible. In accomplishing these aims it is useful to align the camera at near grazing incidence to the x-ray tube target, though there are limitations to this procedure.

Rules for systematically accomplishing the above purposes are derived herein for the extremes of symmetrical (pinhole) and unsymmetrical (slit) aperture systems, as well as for ones employing a simple class of rectangular (unsymmetrical pinhole) aperture. Criteria are also given for evaluating the relative speeds of cameras which may depart from optimum because of mechanical limitations. Finally is described the construction of a series of cameras, found useful in studying the fibrous proteins, particularly collagen, and typical diffraction patterns obtained therefrom are reproduced.

IN recent years there has arisen increasing interest in the use of the diffraction method to study the large structures which are investigated at small diffraction angles. In this development the diffraction method furnishes information which overlaps the size range of objects directly imaged in electron micrographs, and the two methods furnish complementary information about aspects of colloidal structures which could only be indirectly investigated previously.

The small-angle diffraction methods, usually seeking to retain the convenient Cu radiation but to measure structures about a hundred times larger than the common range, have been forced to employ long collimating systems, narrow beam-defining apertures and long specimen-to-film distances. Each of these technical refinements decreases the radiation available at the registering film, so that photographic exposures become matters of days, even of weeks. Consequently, the small-angle diffraction investigator becomes exceedingly conscious of the necessity of improving the intensity of his initial x-ray source or of using it effectively.

Investigators have attempted to solve the peculiar requirements of this field by use of longer wavelengths to increase diffraction angles[1–3] crystals for monochromatization and collimation,[4–7] high intensity

* This paper represents a partial report on research sponsored by the Office of the Quartermaster General, Research and Development Branch, under Project No. 130-46 on "Determination of the nature and properties of skin structure," under direction of the Leather Subcommittee of the National Research Council Committee on Quartermaster Problems.

** Present address: PC Division, Physics Branch, Camp Detrick, Frederick, Maryland.

[1] G. L. Clark and K. E. Corrigan, Ind. Eng. Chem. **23**, 815 (1931).
[2] R. Hosemann, Zeits. f. Physik **114**, 133 (1939).
[3] K. L. Youdowitch, J. App. Phys. **20**, 174 (1949).
[4] I. Fankuchen and M. H. Jellinek, Phys. Rev. **67**, 201 (1945).
[5] J. W. M. Du Mond, Phys. Rev. **72**, 83 (1947).
[6] W. W. Beeman and P. Kaesberg, Phys. Rev. **72**, 512 (1947).
[7] Kaesberg, Ritland, and Beeman, Phys. Rev. **74**, 71 (1948).

x-ray tubes[8,9] or cameras with focusing by means of curves crystals[9,10] or totally reflecting plates[11] to increase the level of useful radiation, and Geiger counters[12] or other x-ray sensitive devices for registration of the diffraction effects.

Each variation has its special advantages and disadvantages. Focusing cameras have improved the situation with respect to slit collimation only, and have not often been used to secure pinhole beams. A similar defect exists for counter registration of diffraction, which becomes difficult if an area rather than a linear diffraction field is to be surveyed. The conventional system, in which the beam is defined by apertures and the pattern registered on photographic film, has many features which recommend themselves. These are, for example, the obtainment of the utmost simplicity of apparatus, the securing of a permanent record of all detail within a given diffraction field, and the integrating properties of the sensitive emulsion, obviating the necessity of exceedingly stable x-ray sources in intensity measurements.

Despite the common use of apertures, there exists no satisfactory consideration of the proper way to employ such collimating systems for maximum camera speed. Hosemann[2] derived optimum dimensions for a collimator, but made the unnecessary assumption that both defining apertures be the same. Yudowitch[3] considered the problem of obtaining apertures of the proper geometry to reproduce details of diffuse small-angle scatter faithfully. In the present paper criteria useful in designing small-angle cameras to obtain adequate small-angle resolution in minimum time are developed. Application is made in the description of cameras successfully employed in studying protein fibers.

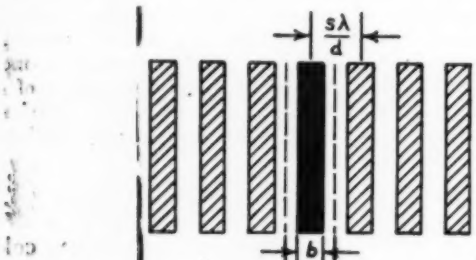


FIG. 1. Diagram of the consecutive diffraction orders of low index yielded at small diffraction angles by a material such as collagen, when examined with a slit camera. The solid bar of width b is the cross section of the undiffracted beam, while the hatched bars of the same width represent the diffraction "images" of the central beam. In order to photograph the diffractions the area of the film outside of the dotted limits, of width a, must be protected from aperture spill-over. The first-order diffraction radius is $s\lambda/d$, where d is the fundamental spacing of the substance examined, s is the specimen-to-film distance of the camera, and λ is the wave-length of x-rays employed.

TYPES OF SMALL-ANGLE INVESTIGATION

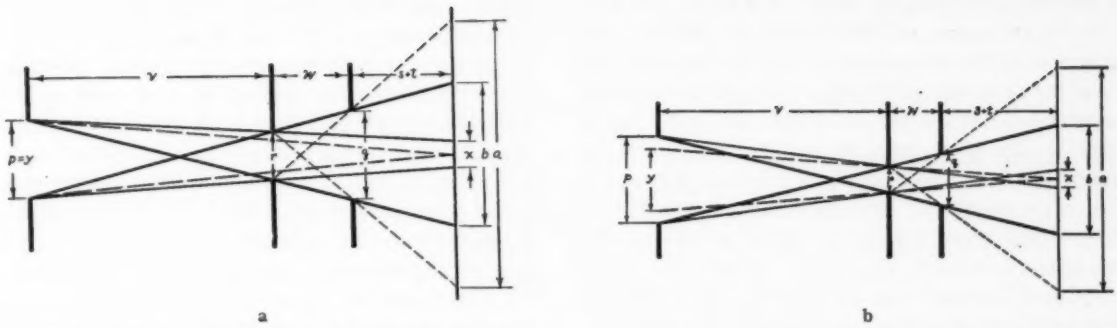
Condensed systems yield distributions of scattered or diffracted radiation at small angles which fall into two distinct classes:

(1) In one class scattered intensity is a continuous function of diffraction angle. From a quantitative evaluation of the intensity distribution, information regarding the occurrence and distribution of irregular structural discontinuities (such as particle boundaries in a colloid, or fibril edges in a fibrous system) may be obtained. Because of its rather general applicability to colloidal systems, the study of this diffuse scatter has been the most prevalent variety of small-angle investigation, and summaries of the methods, theory and results are available.[13–15]

It is generally agreed that studies of diffuse scatter require beam monochromatization, so that a high intensity source and a curved-crystal focusing camera are desirable. The considerations of the present paper would be useful in such work only in attempts to observe pinhole scatter by the use of apertures to collimate the direction of the beam not collimated by the crystals.

(2) A second class of small-angle investigation obtains diffractions consisting of more or less sharp intensity maxima, to which Bragg-law or similar calculations seem applicable and yield large repetitive spacings. Many of these are greater than about 100A, several examples being the following: Gels of tobacco mosaic virus exhibit spacings as large as 280A.[16] An identity period of 171 to 184A has been observed in the protein and lipid complex of the vertebrate nerve myelin sheath.[17] A number of native protein fibers, among them collagen[7,18,19] and keratin and muscle fibers[8,20] show fiber periods ranging from 95 to 725A, some of which (collagen[21–23] and paramyosin[24]) have been di-

[8] I. MacArthur, Nature **152**, 38 (1943).
[9] A. Guinier and G. Fournet, paper presented at New Haven meeting of the American Society for X-Ray and Electron Diffraction, April 1, 1948. Abstract in Am. Mineral. **33**, 759 (1948).
[10] A. Guinier, Ann. Phys. **12**, 161 (1939); Comptes Rendus **223**, 31 (1946).
[11] W. Ehrenberg, Nature **160**, 330 (1947).
[12] Jellinek, Solomon, and Fankuchen, Ind. Eng. Chem. Anal. Ed. **18**, 172 (1946).
[13] I. Fankuchen and H. Mark, J. App. Phys. **15**, 364 (1944).
[14] W. Nowacki, Schweiz. Chem.-Ztg. No. 14–15, 7 (1946).
[15] C. G. Shull and L. C. Roess, J. App. Phys. **18**, 295 (1947).
[16] J. D. Bernal and I. Fankuchen, J. Gen. Physiol **25**, 111 (1941).
[17] Schmitt, Bear, and Palmer, J. Cell. and Comp. Physiol. **18**, 31 (1941).
[18] R. S. Bear, J. Am. Chem. Soc. **64**, 727 (1942); *ibid.* **66**, 1297 (1944).
[19] O. Kratky and A. Sekora, J. makromol. Chem. **1**, 113 (1943); O. Kratky, Monats. f. Chem. **77**, 224 (1946); J. Polymer Sci. **3**, 195 (1948).
[20] R. S. Bear, J. Am. Chem. Soc. **66**, 2043 (1944); *ibid.* **67**, 1625 (1945).
[21] Hall, Jakus, and Schmitt, J. Am. Chem. Soc. **64**, 1234 (1942); Schmitt, Hall, and Jakus, J. Cell. and Comp. Physiol. **20**, 11 (1942); F. O. Schmitt and J. Gross, J. Am. Leather Chemists' Assoc. **43**, 658 (1948); F. O. Gross and J. Schmitt, J. Exp. Med. **88**, 555 (1948).
[22] C. Wolpers, Klin. Wochschr. **22**, 624 (1943); *ibid.* **23**, 169 (1944); Virchow's Arch. Path. Anat. **312**, 292 (1944).
[23] G. C. Nutting and R. Borasky, J. Am. Leather Chemists' Assoc. **43**, 96 (1948).
[24] Schmitt, Hall, and Jakus, J. App. Phys. **16**, 263 (1945); Schmitt, Bear, Hall, and Jakus, Ann. N. Y. Acad. Sci. **47**, 799 (1947).



a b

FIG. 2. Illustrating the geometry of collimating systems, with aperture and beam cross sections magnified relative to the various separations between the apertures and the film. Figure 2a shows an arrangement in which the film dimension x, over which is possible maximum view of the front aperture, is defined as *positive*, while Fig. 2b indicates the situation for *negative* x. In both cases y is the actual width viewed, from the film center, of the total front aperture p. Other apertures are r, the principal or rear defining aperture, and q, the guard aperture. At the film, b and a are the corresponding quantities of Fig. 1. The collimator lengths are v and w as indicated, while the specimen (not shown) is placed between the guard aperture and the film at a distance t from the former and s ahead of the latter.

rectly photographed in the electron microscope. The fibrous mineral crysotiles[25] and the synthetic polyamides[13] have also been reported to diffract moderately sharply at small angles.

With Cu radiation the first-order Bragg angles of spacing between 100 and 1000A, in which range the above examples lie, are from 0.5 degree to about 3 minutes, and twice these values are the separations of the corresponding diffractions from the undeviated beam. In arranging for the resolution of these very small angles, as follows below, one obtains about all the beam refinement that is likely to be required for investigations profitably using the diffraction method.

CAMERA REQUIREMENTS

In considering the problem of obtaining a camera capable of examining the consecutive diffraction orders of a single fundamental spacing, the investigator has in mind essentially three requirements: (1) *angular resolution*, which involves separation of the diffraction effects from each other and the central undeviated beam; (2) *magnification*, permitting sufficient distance between diffraction line centers to allow adequate accuracy of spacing measurement; and (3) maximum *camera speed* for rapid registration of the diffraction pattern. The first two of these requirements have been considered briefly in another place,[18] while the third is the chief concern of the present paper.

The desired angular resolution and magnification initially impose certain limitations on the intensities obtainable, so that consideration of the first two factors must precede examination of the third. The resolution desired has two aspects: separation of consecutive orders, whether low or high, and guarding of beam spill-over from defining apertures so that low orders can be seen clearly. Figure 1 illustrates the situation.

Assume that each diffraction order very nearly reproduces the width, b, of the undiffracted beam at the film (see discussion below of unsymmetrical collimating systems). At the small diffraction angles in question, where sines and tangents become equal to their arguments, the distance between the consecutive orders becomes $s\lambda/d$, where s is the specimen-film separation, λ the wave-length employed, and d the fundamental spacing of the material. The resolution of consecutive orders can be accomplished if

$$b = s\lambda/fd, \tag{1}$$

when f is a factor equal to unity or greater. The product fd, expressing the maximum spacing whose orders are resolvable by a given camera, is a figure of merit for the camera's small-angle resolution.

For the g^{th} order (and all higher ones) to be free of beam spill-over, the value of the width of the unguarded region, a, must be equal to no more than the diffraction diameter of the g^{th} order ($2gs\lambda/d$), less the width of a diffraction, b, so that according to Eq. (1),

$$a = (2fg - 1)b, \tag{2}$$

will accomplish the desired guarding. Since it is physically impossible for a to be less than b, it follows that g must always be greater than $1/f$. The factor g is a convenient expression of a camera's ability to guard the diffractions corresponding to a fundamental spacing d, stated in terms of the index of the lowest order that remains guarded, though g need not be limited to integral values.

Figure 1 shows a desirable choice of the relative widths of central and diffracted beams and of unguarded area at the film, the latter being covered normally by a lead beam stop slightly greater than a. The values of f and g illustrated are 2 and 0.75, respectively, resulting in separation of adjacent orders by a clear space equal to their individual widths (b), while guarding ($a = 2b$) bisects the spaces between the incident beam and the first orders.

The ratio s/d offers a convenient expression for the magnification of the diffraction pattern relative to the

[25] I. Fankuchen and M. Schneider, J. Am. Chem. Soc. **66**, 500 (1944).

size of fundamental spacing. With this ratio held constant, for examination of substances of different d's, the pattern of diffraction orders at the film is independent of d as far as positional effects for diffraction maxima are concerned. Then the factors of film grain, shrinkage, etc., affecting the percentage accuracy of measurement, are independent of spacing. It is obvious physically that the ratio should be held as small as possible for the most rapid photographic registration of the diffraction pattern, though the ratio cannot be indefinitely reduced because the collimating apertures become too small for accurate construction and control.

In many laboratories the ratio s/d (expressed as mm/A) is chosen 0.5 or greater, because it is found more convenient to construct such cameras and view the resulting patterns. In a previous publication,[18] for example, 300-mm specimen-to-film distances were used to study the macroperiod (*ca.* 600A) of collagen fibers. Subsequently it has been feasible to diminish the ratio to 0.25 or lower by reducing the specimen-to-film distance to 150 mm or less, without too great deterioration in pattern quality. On occasion, cameras with 0.125 ratio are possible as far as the obtaining of consecutive diffraction order resolution is concerned, but low ratios in this range can best be used when low orders can be sacrificed. Patterns obtained at low s/d ratios may be described as micro-diagrams, since they require a small optical magnification to be observed conveniently by eye.

The above paragraphs express as quantitatively as possible the matters involving judgment in deciding the specifications of a small-angle diffraction camera. Obviously, to investigate a substance with fundamental spacing d, most rapid results are obtained at the smallest magnification ratio (s/d), the smallest f factor, and the largest g consistent with other requirements. With these three factors decided, one has thereby fixed, with the help of Eqs. (1) and (2), the camera dimensions s, b, and a.

THE COLLIMATION GEOMETRY

Sometimes it becomes desirable to obtain a camera capable of resolution in more than the one film direction hitherto considered. Then the magnification is governed by the largest d to be examined, but the resolution and guarding needed for the two orthogonal directions on the film, as expressed in b and a pairs, are independent for the two directions. If both pairs of b and a are the same, apertures will be round or square ("pinhole" cameras), but if different, then rectangular or slit apertures are useful in order that no more limitations be placed on the openings of the camera than are necessary. In what follows vertical and horizontal collimations are distinguished, the vertical collimation being that for resolution in a corresponding (meridional, subscript 1) direction on the diffraction pattern, the horizontal direction being for the remaining orthogonal (equatorial, subscript 2) orientation on the film. In present instances highest angular resolution is for vertical orientation, which determines s for both orientations.

There are many ways to vary three apertures and their separations to obtain the a, b, and s which are initially required for a given direction on the film. Figure 2 shows two of these, as well as serving to define terms indicating camera dimensions. Subscripts are appended to these symbols as mentioned above when orientations of collimation need to be distinguished. Since in present cameras the corresponding apertures, the specimen, and the film are to be placed at the same locations along the x-ray beam for both collimation orientations, it is never necessary to use subscripts on s, t, v, and w.

The geometric relationships between camera dimensions are most easily expressed as follows:

$(p+r)/v=(q-r)/w=(b-q)/(s+t)$ (main collimation),
$(p-r)/v=(r-x)/(s+t+w)$ (lines to x limits),
$(r+q)/w=(a-q)/(s+t)$ (guard collimation).

For convenience these may be rearranged to

$$\left.\begin{aligned} r&=\frac{(b+x)p}{(b+2p-x)},\\ q&=\frac{(a+b)(b+x)p}{(a-b)(b-x)+2(a+x)p},\\ v&=\frac{2(s+t)[(a-b)(b-x)+2(a+x)p]p}{(a-b)(b-x)(b+2p-x)},\\ w&=\frac{2(s+t)(b+x)p}{(a-b)(b+2p-x)}. \end{aligned}\right\} \tag{3}$$

As here expressed, one may regard p and x as independent variables, and a, b, and s as predetermined constants. The variable t is a trivial one, which should be as small as possible for maximum camera speed and will be determined by experimental factors involving specimen manipulation, often being negligible. Two such sets of equations can be written, one set for each of the two collimation orientations.

THE INTENSITY PROBLEM

In deciding how to choose p and x one must obtain an intensity expression which will state the function of p and x to be maximized for greatest camera speed. It is desirable that the center of the undiffracted beam at the film receive radiation energy at a maximum rate. As reference to Fig. 2 indicates, the film center receives radiation from an area y_1y_2 of the front aperture at a distance therefrom of $s+t+v+w$. Consequently, energy is received at the film center at a rate proportional to the function $y_1y_2B/(s+t+v+w)^2$, where B is the energy flux per unit area available at the front aperture. The assumption is made that no absorption occurs in the camera, which must therefore be evacuated.

It is through the factor B that the suitability of a

given x-ray tube for diffraction cameras enters, particularly in small-angle studies. The energy collected by a small film area from the entire target, which is well known to be best at angles of 4° and 6° or more from glancing incidence on the target face, is not involved here. The desirable quality is obtained much nearer to grazing incidence.

To investigate the factor B for a General Electric CA-6 Cu target tube, with which the present work was done, a collimator was constructed having two horizontal slits (corresponding to the vertical collimation described above), the one nearest the tube having dimensions 0.05×1.0 mm and the second 0.1×1.0 mm. The tube target face was horizontal, and the collimator was aligned at various angles to the plane of the tube target, looking in the direction of the line focus. The direct beam was then photographed at various incidences, and the density of the center of each line determined with a microphotometer. Plotted against glancing angle, α, between collimator axis and target face, the results shown in Fig. 3 indicate how B is influenced by change in α, since at all times the projected focal spot was larger than the collimator's front aperture, which remained of constant area and distance from the film.

Above angles greater than about 0.5° the factor B is roughly proportional to $\csc\alpha$. B is then proportional to the total area of the focal spot which can send radiation to the film line center. This would be expected only if surface target atoms are emitting x-rays symmetrically. Actually the intensity decreases at extremely low angles, indicating that the average source is not exactly at the surface layer but some distance below. The projected focal spot "length" (which actually becomes less than the focal spot width at these grazing angles) for which the factor B is a maximum is hereinafter designated as p_0; in the CA-6 tube this is about 0.1 mm.

An important item in the construction of tubes for small-angle work is the requirement that tube windows allow collimators to be directed toward the target at near grazing incidence. Guinier[26] has also reported a gain in intensity obtained by viewing the focal spot in this manner.

Since the line focus on the target is limited in length and remains constant for all angles of incidence, α, at grazing angles its available projection normal to the axis of the collimating system (hereafter termed p_m) is proportional to α. The front aperture of the collimator is close enough to the target so that this projection (p_m) determines the widest front aperture which can be used at a given α. According to the relations between B and α given above, it follows that at grazing angles B is inversely proportional to p_m. The variable p_m is independently controllable by varying α, so must also be chosen to maximize the intensity received at the film center. One may therefore adopt the function,

$$I=\frac{y_1y_2}{(s+t+v+w)^2p_m}, \tag{4}$$

as the one determining the intensity applied by tube and camera to the film center.

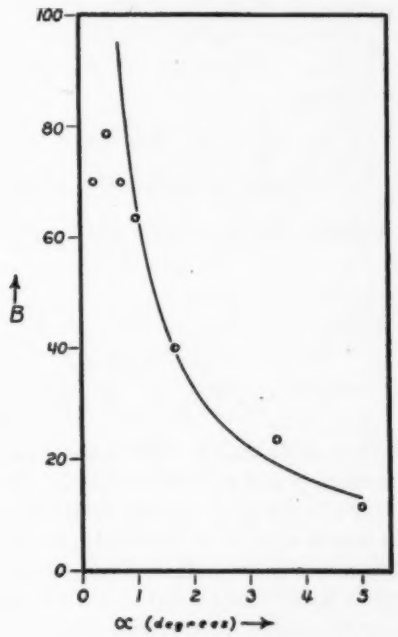


FIG. 3. The relation between B, a factor expressing the available radiation flux per unit area at the front aperture of a collimating system, and α, the glancing angle between the line focus of the tube target and the axis of the collimator. B is given in arbitrary units, α in degrees. Circles are experimental points obtained from a General Electric CA-6 copper target tube, while the curve shows the behavior expected if B is proportional to $\csc\alpha$.

The variable p_m applies only to vertical collimation when the tube target plane is horizontal; the width of the focal line requires the aperture p_2 to remain within a single maximum value which remains constant at all angles of incidence, since foreshortening of the target is possible for only one direction. The present discussion utilizes vertical collimation orientation for resolution of largest spacings because of this fact, in anticipation of use with tubes having horizontal focal planes and because in this manner one can conserve target energy best for the most difficult resolution problem involved in a given instance. If a tube is oriented with target plane vertical, interchange the vertical and horizontal relationships expressed for camera and film orientations herein.

One can simplify the problem of maximizing I by considering the two most directly useful cases: that in which vertical and horizontal resolution and guarding are equal (symmetrical pinhole case) and the extreme unsymmetrical case wherein no resolution and guarding are required of the horizontal collimation (slit camera).

[26] A. Guinier, *Radiocrystallographie* (Dunod, Paris, 1945), p. 80.

SYMMETRICAL PINHOLE CAMERAS

In the symmetrical pinhole camera, $I=y^2/(s+t+v+w)^2p_m$, where no additional subscripts are necessary, since the corresponding quantities (except for p_m) are alike for the two collimations. It is necessary to distinguish further the two situations with regard to positive and negative x shown in Fig. 2, since these determine y differently. When x is positive, $y=p$, which with relations expressed in Eqs. (3) yields

$$I_+=\frac{p^2}{(s+t+v+w)^2p_m}=\left\{\frac{(a-b)(b-x)p}{(s+t)[(a-b)(b-x)+2(a+x)p]}\right\}^2\frac{1}{p_m}.$$

When x is negative, $y=r(s+t+v+w)/(s+t+w)$ and

$$I_-=\frac{r^2}{(s+t+w)^2p_m}=\left\{\frac{(a-b)(b+x)p}{(s+t)[(a-b)(b-x)+2(a+x)p]}\right\}^2\frac{1}{p_m}.$$

Differentiation of I_+ and I_- with respect to the variable x, at constant p and p_m, shows that $\delta I_+/\delta x$ is negative and $\delta I_-/\delta x$ is positive for all realizable systems (i.e., ones in which $a>b>|x|$, and all quantities other than x are always positive). It is clear, therefore, that greatest camera speed is obtained at $x=0$ in both positive and negative ranges for this variable.

In a similar way, the partial derivative of I (subscripts may be dropped now that $x=0$) with respect to p at constant p_m show that for all realizable cameras $\delta I/\delta p$ is positive, or that the front aperture should be as wide open as possible. One cannot, however, enlarge the aperture farther than p_m usefully.

At this point either pinhole intensity expression becomes (letting $x=0$, $p=p_m$)

$$I_P=\left\{\frac{(a-b)b}{(s+t)[(a-b)b+2ap_m]}\right\}^2p_m, \qquad (4a)$$

which has a maximum at $p_m=(a-b)b/2a$. If the value of p_m calculated from this expression is less than the minimum practical projected focal spot dimension, p_0, available at the tube, one uses $p=p_m=p_0$.

SLIT COLLIMATION

In slit cameras one permits the front aperture, p_2, for horizontal collimation to remain constant at the invariable value approximating the width of the target focal spot. Then $I=y_1p_2/(s+t+v+w)^2p_m$, since no other apertures are to limit the target view of the film center in horizontal collimation. Only the set of Eqs. (3) with subscript 1 are needed, though this subscript can be omitted without confusion.

When x is positive,

$$I_+=\left\{\frac{(a-b)(b-x)}{(s+t)[(a-b)(b-x)+2(a+x)p]}\right\}^2\frac{p_2p}{p_m}.$$

As in the symmetrical pinhole case it can be shown that $\delta I_+/\delta x$ is always negative. Consequently, for positive x the choice $x=0$ is best, throwing the determination of optimum conditions into the range of negative x, in which

$$I_-=\left\{\frac{(a-b)}{(s+t)[(a-b)(b-x)+2(a+x)p]}\right\}^2\frac{(b^2-x^2)p_2p}{p_m}.$$

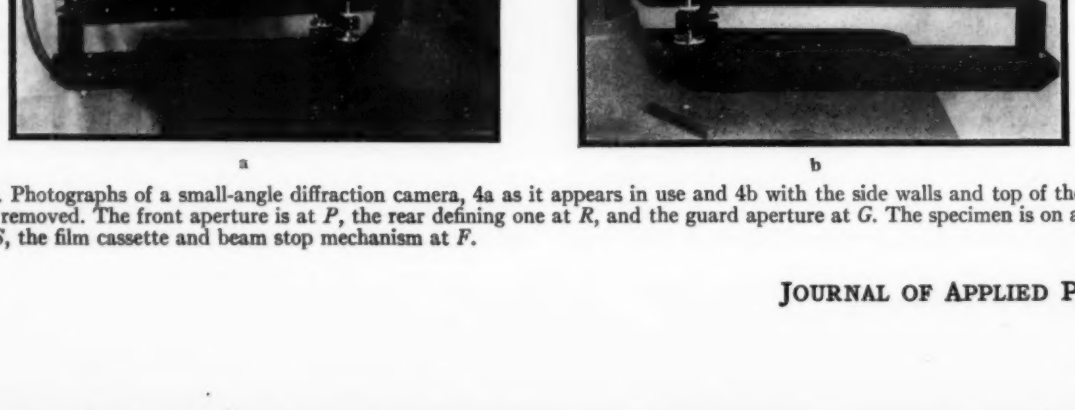
For any realizable magnitude of x, the optimum value of p is given by $p=(a-b)(b-x)/2(a+x)$.

In order to examine conditions with respect to x more clearly it is convenient, however, to let $p=(a-b)\times(b-x)z/2(a+x)$, where z is a new positive variable, replacing p temporarily. Then,

$$I_-=\frac{(a-b)p_2}{2(s+t)^2p_m}\cdot\frac{(b+x)}{(a+x)}\cdot\frac{z}{(1+z)^2},$$

and

$$\frac{\delta I_-}{\delta x}=\frac{(a-b)^2p_2}{2(s+t)^2p_m}\cdot\frac{1}{(a+x)^2}\cdot\frac{z}{(1+z)^2}.$$



a b

FIG. 4. Photographs of a small-angle diffraction camera, 4a as it appears in use and 4b with the side walls and top of the vacuum chamber removed. The front aperture is at P, the rear defining one at R, and the guard aperture at G. The specimen is on a separate rider at S, the film cassette and beam stop mechanism at F.

It is readily apparent, therefore, that for realizable cameras (i.e., $a>b>|x|$) $\delta I_-/\delta x$ is always positive, hence optimum x is always zero.

The possibility that p may be less than $(a-b)b/2a$ is retained because the maximum value that p can have is p_m, which could be less than this optimum. It remains to be determined whether, when $p=p_m$, anything is to be gained by letting p_m become smaller, since it is obvious that larger p_m is not indicated. Letting $p=p_m$, one finds that

$$I=\left\{\frac{(a-b)b}{(s+t)[(a-b)b+2ap_m]}\right\}^2 p_2,$$

so that $\delta I/\delta p_m$ is always negative. Consequently, in any event p_m equals p_0, its minimum practical value, and p assumes its optimum value $(a-b)b/2a$ only if this is smaller than p_0, otherwise being set at p_0.

Because of the fact that p may not always be equal to p_m, an intensity expression which is generally valid for slit cameras, requiring only that $x=0$, must be obtained from the original expressions for I_+ and I_-:

$$I_S=\left\{\frac{(a-b)b}{(s+t)[(a-b)b+2ap]}\right\}^2 \frac{p_2 p}{p_m}. \qquad (4b)$$

THE UNSYMMETRICAL PINHOLE CAMERA

If an unsymmetrical collimating system must yield two different angular resolutions with unequal guarding, it will be found that the most general problem of determining the optimum camera becomes complex. There is, however, a circumstance in which the unsymmetrical pinhole case becomes no more difficult than the symmetrical one.

Enlarge all apertures, the beam stop, and beam widths of the horizontal collimation (subscript 2) over the corresponding dimensions of the vertical system (subscript 1) by multiplication of the latter by a common factor j. It is found from Eq. (1) that $(f_1d_1/f_2d_2)=j$ and from Eq. (2) $f_1g_1=f_2g_2$, since s must be the same for both collimations. These two relations show the latitude allowed in f and g for the two collimations when a material with orthogonal spacings d_1 and d_2 is to be examined with a given camera.

Equations (3) show that this treatment is consistent for such dependent apertures as r and q, while v, w, s, and t of both collimations remain the same. The intensity relationship now becomes the same as in the previous symmetrical pinhole case, except for multiplication by j when the variables for vertical collimation are employed. Choice of optimum camera dimensions proceeds as in the symmetrical case, leading to the optimum x_1, p_1, and p_m, from which r_1, q_1, v, and w may be calculated. The subscripted quantities are then multiplied by j to arrive at the corresponding apertures of the second collimation, except that there is no second p_m involved. In most instances the target focal spot is so wide that its limitation of p_2 is not experienced.

This method may fail to gain all possible advantage from the asymmetry of the tube target or the minimum guarding conditions actually required by a given problem. However, it controls resolution, guarding, and magnification and will produce cameras which are faster than the symmetrical ones designed to handle the higher spacing, f_1d_1, in both directions.

The gain in speed attending use of unsymmetrical cameras, or indeed more generally realized from any means of employing as wide open apertures as possible, depends not only on camera dimensions but also on certain qualities of the object. The latter must be capable of a small degree of diffraction spread, so that at a given spot on the film diffracted rays, produced by corresponding components of the incident beam passing through different portions of the object at slightly different angles of incidence, may cooperate in building up the registered density. In this way, increase of the amount of radiation passing through the object results in registration of greater sized diffractions and greater density within each diffraction.

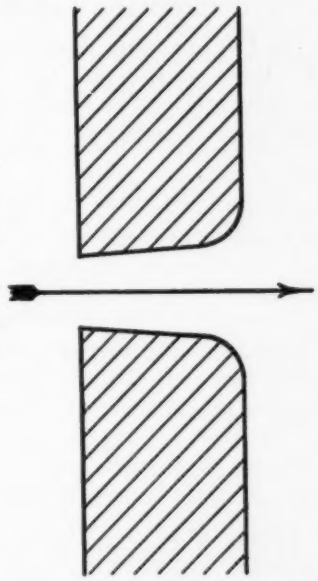


FIG. 5. Cross-sectional shape of the lead pieces which are adjusted under a microscope to the desired separation in forming an aperture. The arrow shows the direction in which the x-ray beam passes the aperture.

The spreading of diffraction may result from several specimen properties. Most obvious cause is a lack of perfect orientation of the small diffracting elements (individual crystals, fibrils, etc.) of the sample, as in a powder, for which the use of slit collimators has long been recognized to be of advantage. Less apparent are the intrinsic structural causes for diffraction diffuseness: small size or imperfection of order along one or more of the dimensions of the diffracting elements. At small angles subjects of study probably never lack all the causes of diffraction spread just described, so that in this diffraction field the present methods of setting up resolution conditions and criteria for camera speed seem of fairly general application.

TABLE I. Dimensions of typical small-angle cameras.

j	fd	s	a	b	p	q	r	t	v	w	I_R	f	g
∞	1400	150	0.306	0.165	0.073	0.084	0.039	10.0	220	88	0.90	2.2	0.65
1.0	1200	100	0.256	0.128	0.100	0.072	0.039	2.0	255	61	1.00	1.9	0.80
2.0	800	150	0.680	0.289	0.172	0.138	0.078	3.0	255	61	0.89	1.25	1.3
∞	600	100	0.514	0.257	0.100	0.117	0.056	3.0	115	45	0.95	0.94	1.6
2.0	400	150	1.12	0.578	0.256	0.293	0.136	10.0	220	88	0.91	0.63	2.3

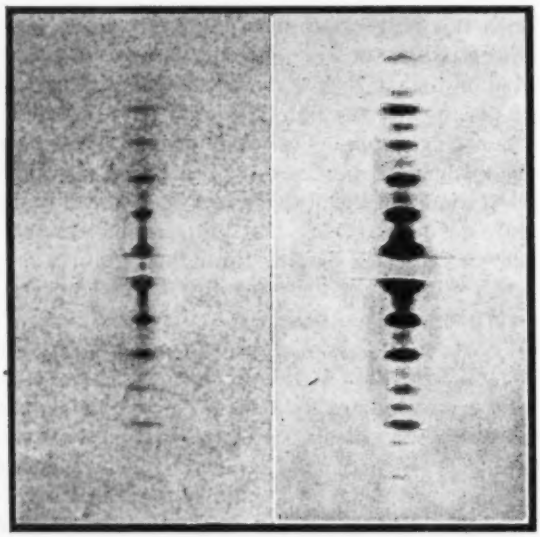
Explanation: The symbols used are those of Fig. 2 or explained relative to Eqs. (1) and (2), with fd in A and columns from s through w in mm. Except for s, t, v, and w, the dimensions and resolutions given are for vertical collimation only, the factor j being used in multiplying dimensions and to divide resolution in going to the horizontal direction, with ∞ denoting a slit camera. The factors f and g depend on the d of each specimen examined, and are given here for dry collagen (d=640A). I_R represents the intensity theoretically expected of each camera (all have x=0) relative to that of the optimum one required for each given task, assuming an x-ray source with p_0=0.1 mm.

CHOOSING CAMERA DIMENSIONS

In the preceding discussion it has been shown that for both pinhole and slit cameras, the independent variable x should be zero. Consequently, in actual practice Eqs. (3) reduce to

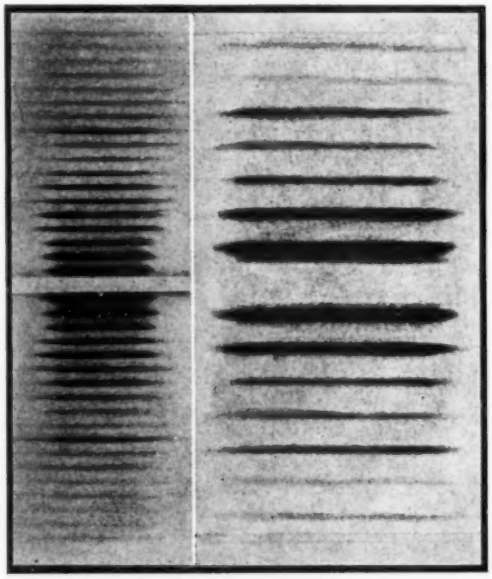
$$\left.\begin{aligned} r&=\frac{bp}{(b+2p)},\\ q&=\frac{(a+b)bp}{(a-b)b+2ap},\\ v&=\frac{2(s+t)[(a-b)b+2ap]p}{(a-b)b(b+2p)},\\ w&=\frac{2(s+t)bp}{(a-b)(b+2p)}. \end{aligned}\right\} \tag{3a}$$

On the other hand, the choice of the remaining independent variables, p and p_m, is different for pinhole and slit cameras. One first calculates a *tentative* forward aperture, $p'=(a-b)b/2a$. This is the optimum choice for p, provided p' bears a certain relationship to p_0. If p' is less than p_0, it is inadvisable to let $p=p'$ for *pinhole* cameras, since in so doing excessive target area will be lost when the camera is aimed at the tube. On the other hand, with *slit* cameras all values of p greater than p_0 are undesirable, since they do not increase useful target area and only increase collimator length. The projected focal line length (p_m) of the (horizontal) tube target must be adjusted in aligning the camera so that p_m just fills the front aperture for the vertical collimation, though p_m must never be less than p_0. With these facts in mind, the criteria for selection of p and p_m may be summarized as follows: In pinhole cameras p always should equal p_m, both being equal to p_0 or p', whichever is the greater. In slit cameras, p_m always should equal p_0, but p should equal p_0 or p', whichever is the smaller.



a b

FIG. 6. A comparison of the small-angle collagen patterns obtained with the 1200 (6a) and 800 (6b) pinhole cameras of Table I from kangaroo tail tendon specimens which were kept moist during exposure. The characteristic alternation of intensities (odd orders strong) and relative constancy of line lengths laterally, typical of wet specimens, is apparent. These photographs cover diffraction angles under about 1.5°. In this and Fig. 7 the fiber axis is always vertical.



a b

FIG. 7. Photographs of collagen small-angle diffractions obtained with the 1400A slit camera of Table I. The area of the first twenty orders is shown in 7a, while 7b shows the central portion of the same diagram at much higher magnification as it appears on a second film in the cassette, on which the resolution of the first order is clearly accomplished. This specimen has been "stained" with phosphotungstic acid to the point where the normal intensity relations of the orders are changed.

There are often, however, practical reasons which make it desirable to depart from optimum conditions. For example, the best magnitude of the aperture r may be too small for convenient fashioning, or a number of available camera parts (such as a collimator of fixed length) may make it inconvenient to attain the best situation. While the most economical variation from optimum x, p, or p_m depends on the particular situation, the following general procedure has proved useful in designing cameras applied to the extreme low angles necessary for protein fibers. It is assumed that x should always be zero, and that p_m equals p except in the rare occasion (in slit cameras) when the latter is reduced below p_0, which is always the minimum value for p_m. Under these conditions the previously developed Eqs. (4a) and (4b) may be used to express camera intensities even when optimum p and p_m are not chosen.

One may calculate intensities for both optimum and actual cameras, in both of which the same a, b, s, and t express the desired result. It may be thus determined that choice of certain p and p_m, selected to permit the aperture r or the total collimator length to reach desired values, does not cause too great intensity loss. A convenient way to express the result is that of a relative intensity, I_R, which is the ratio of the intensity of the actual camera to that of the hypothetical optimum one.

A general formula applicable to both pinhole and slit cameras described herein is:

$$I_R=\frac{p(p'+p'')^2}{p''(p'+p)^2}, \tag{5}$$

where p is the actual front aperture employed, p' is the tentative optimum aperture, and p'' is the real optimum value dictated by the rules which have been given. This equation assumes for pinhole cameras that the actual p and p_m used are equal and never less than p_0. It also assumes for slit cameras that the actual p_m equals p_0 when $p\leq p_0$. When $p>p_0$ it is necessary also for the p_m employed to be equal to p, in which case an additional factor p_0/p_m must be used. The last case does not arise for the present instances, but might be necessary for slit cameras with smaller spacing resolution (fd) than is met with in Table I.

An expression useful in considering cameras in which $p\neq p'$ follows from Eqs. (3a), which show that

$$v+w=\frac{p}{p'}(s+t), \tag{6}$$

indicating the way in which collimator length is affected by variations in front aperture, p.

Equations (3a), (4a), (4b), (5), and (6) should never be employed for cameras in which $x\neq 0$. For example, the common type of system with equal defining apertures ($p=r=x$), which intensity calculations indicate to be distinctly undesirable, requires a return to first principles (Eqs. (3) and (4)).

During this investigation cameras departing widely from optimum have been constructed. *Quantitative* comparison of these cameras with the nearly optimum ones herein described, along the lines of the present camera-speed criteria, is difficult to achieve. Nevertheless, experience has at many points checked *qualitatively* the conclusions of these criteria, yielding a considerable degree of confidence in their usefulness.

TYPICAL SMALL-ANGLE CAMERAS

Figure 4 shows side views of a typical small-angle camera employed in the study of the fibrous proteins. The box cover carries a rubber gasket which seals the system for evacuation. The side walls of the vacuum chamber are sealed with wax to the base and connected by means of thick rubber tubing to the collimating tube, since solid connections tend to transmit strains and distortions of the walls, caused by the evacuation, to the base and tube which bear the critically aligned elements. The collimating tube is collapsible to permit extending the front aperture toward the tube window without movement of the whole camera, the joint being sealed with wax in use. The front aperture is closed with the Ni filter employed for partial monochromatization. The entire camera is mounted on a 3-in. channel iron supported by three graduated leveling screws.

Each aperture and the specimen are placed on separate inserts, so that different choices of aperture may be made and specimens readily changed. Separate riders bear the second or main defining aperture, the guard aperture, the specimen, and the film cassette holder with its adjustable beam stop. Each rider is firmly clamped to the base in use. In aligning the camera, the front and main defining apertures, in their fixed positions, establish the beam line, to which the guard aperture and beam stop must be adjusted by means of screws acting against springs. Once the camera elements themselves are in order, the entire camera is aimed at the tube target by means of the external leveling screws until the fore-shortened focal spot just fills the vertical dimension of the front aperture (unless $p<p_0$). Critical final adjustments are followed photographically.

The apertures are made as follows: a flat piece of lead approximately 0.025-in. thick, and of suitable dimensions otherwise to fit the eventual aperture insert, is cut along a middle line with a sharp razor blade. The edges are lightly ground on an Arkansas stone and finally polished by rubbing on paper to give a smooth edge of sectional shape shown in Fig. 5. It is important to have the defining edges sharp and placed toward the incident beam in order to minimize the area capable of total reflection toward the film.[27] The lead pieces are

[27] The chief effects which guarding must eliminate are total reflection from the edges of collimating apertures and Bragg reflections from the crystalline material forming the apertures. At small angles the latter are not important. Since total reflection is

then cemented to the aperture insert and before the cement has hardened adjusted under a microscope until parallel and of the desired separation. This completes preparation of a slit, and pinholes involve the superposition of a second pair of jaws normal to the first pair, though in some instances separate inserts for vertical and horizontal apertures have been made, making possible rapid conversion of slit into pinhole cameras, and vice versa.

As beam stop, a lead piece, of appropriate shape and size to cover the beam and the adjacent unguarded area, is attached to a stiff wire borne by the adjustable sliding mechanism on the cassette support. If the lead is 0.012 in. to 0.014 in. thick the direct beam will be faintly registered on the film in about 24 hours.

Table I shows the essential dimensions of some cameras actually built and used for studying fibrous proteins. These cameras maintain nearly a constant value of $\frac{3}{2}$ for the product of f and g, corresponding to a general practice in this series of maintaining approximately $a=2b$ (cf. Eq. 2). The fd 1200, 800, and 400A pinhole systems, and the 1400A slit system have been employed principally for various purposes in the study of collagen. While the 400A camera should not be suitable for collagen, it is found in fact that the calculated fd values are conservative and that consecutive orders of larger spacings can actually be separated, though the lowest orders are lost. The 400A pinhole and the 1400A slit cameras allow a 10-mm value of t in order to accommodate a special tilting holder for the specimen, as will be described when appropriate.

While the 1200A camera shows exceedingly fine angular resolution when necessary, as in viewing the innermost diffractions, it is often worth retreating in resolution to the 800A or coarser systems to gain greater pattern density for the registration of fainter pattern details. Figure 6 compares the results of the two best pinhole cameras, obtained with moist collagen. On the other hand, slit cameras are capable of both speed and high resolution, though only along the fiber axis. Figure 7a shows the diagram yielded by the 1400A slit camera from a dry collagen sample "stained" with phosphotungstate. Figure 7b shows, at increased magnification, the central section of 7a as it appears on a second film, placed behind the first one in the cassette in order to reduce intensity and avoid halation for the lower orders.

Further description of the details of the collagen diffraction patterns and their interpretation will be given elsewhere. Although the collagen small-angle diagram consists of a single meridional set of diffraction orders, resolvable with slit cameras, considerable interest attaches to the lengths and shapes of the layer lines. Pinhole patterns have been required for the study of these phenomena in collagen, as well as for the resolution of close-spaced, near-meridional row lines exhibited by the keratin-myosin family of fibrous proteins.[20]

limited to tens of minutes of angular departure from the central beam, one would need no guard ($g=\infty$) to resolve moderately large spacings (*ca.* 100A) if the diffractions are intense and sharp enough to emerge from the relatively weak secondary scattering of aperture edges.

Electron Microscope and Diffraction Study of Metal Crystal Textures by Means of Thin Sections*

R. D. HEIDENREICH
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received March 11, 1949)

Bethe's dynamical theory of electron diffraction in crystals is developed using the approximation of nearly free electrons and Brillouin zones.

The use of Brillouin zones in describing electron diffraction phenomena proves to be illuminating since the energy discontinuity at a zone boundary is a fundamental quantity determining the existence of a Bragg reflection. The perturbation of the energy levels at a corner of a Brillouin zone is briefly discussed and the manner in which forbidden reflections may arise at a corner pointed out. It is concluded that the kinematic theory is inadequate for interpreting electron images of crystalline films.

An electrolytic method for preparing thin metal sections for electron microscopy and diffraction is introduced and its application to the structure of cold-worked aluminum and an aluminum-copper alloy demonstrated. It is concluded that cold-worked aluminum initially consists of small, inhomogeneously strained and disoriented blocks about 200A in size. These blocks are not revealed by etching but would contribute to line broadening in the conventional diffraction experiments. By means of a reorientation of the blocks through a nucleation and growth process, larger disoriented domains about 1–3μ in size found experimentally could be accounted for. It is suggested that such a nucleation and growth reorientation phenomenon is responsible for self-recovery in cold-worked metals.

The formation of $CuAl_2$ precipitate particles is demonstrated with both electron micrographs and diffraction patterns. A fine lamellar structure found in the quenched Al-4 percent Cu alloy is at present unexplained.

INTRODUCTION

IT is intended that this paper fulfill a twofold purpose. One aim is to present a more easily understandable development of the dynamical theory of electron diffraction and to emphasize the necessity of employing the theory in interpreting electron images of crystals. The second purpose is to make known in detail an electrolytic technique for preparing thin metal sections and to demonstrate the useful information that can be obtained from such sections.

The dynamical theory of electron diffraction as initially published by Bethe[1] is difficult to follow and often not clear in the mathematical manipulations involved. A later exposition by MacGillavry[2] suffers from similar difficulties, particularly as regards the application of boundary conditions. Other treatments of the theory found in the literature are in general not as straightforward as would be desirable. Since none of these discussions make direct use of the mathematical and physical approach used in the theory of solids, it is felt that a treatment of the dynamical theory based on well-known perturbation methods and Brillouin zones would be instructive in pointing out the close relation between diffraction and the band theory of solids. In addition to placing the theory on a physically more satisfying basis, a primary aim is to stimulate a greater awareness of the necessity for further developing and applying the theory in electron microscope and diffraction experiments with crystalline materials. This is particularly true in the case of thin sections of crystalline materials as will be seen.

The large cross section for electron scattering exhibited by most solids makes it necessary to employ very thin sections for transmission electron microscope and electron diffraction studies. Methods of preparing sufficiently thin sections (*ca.* 500A thick) are in a crude state of development with considerable disagreement even as to the general principles to be followed. An instrument (high speed microtome) has been described by O'Brien[3] and by Fullam[4] based on the premise that a high velocity knife is more likely to produce thin sections. Results obtained with this instrument indicate that in the case of crystalline solids, at least, deformation of the material during cutting is sufficient to drastically change the crystalline texture.

The experimental work to be described herein resulted from a desire to investigate in more detail changes in structure and orientation brought about by plastic deformation.[5] The method of sectioning developed for this purpose is electrochemical since such a method would be expected to produce far less mechanical disturbance than a cutting process. Although the present technique requires considerable refinement, the results indicate that electrochemical or chemical methods hold promise for sectioning metals.

An indication of the type of electron images that might be obtained from thin metal sections can be found in the x-ray "diffraction microscope" experiments of Barrett[6] and of Smoluchowski.[7] In these experiments, the intensity distribution is determined by Bragg reflections from properly oriented regions of the crystal and hence constitutes an orientation map. The resolution is about 10 microns. The effect of cold-work, precipitation etc. is quite noticeable in the x-ray results.

The simplest situation in the case of the transmission electron microscope and a thin section of polycrystalline

* Portions of the experimental part of this paper presented at the Annual Meeting, Electron Microscope Society of America, Toronto, Canada, September 9, 10, 11 (1948).

[1] H. A. Bethe, Ann. d. Physik **87**, 55 (1928).

[2] C. H. MacGillavry, Physica **7**, 334 (1940).

[3] H. C. O'Brien and G. M. McKinley, Science **98**, 455 (1943).

[4] E. F. Fullam and A. E. Geisler, Rev. Sci. Inst. **17**, 23 (1946).

[5] R. D. Heidenreich and W. Shockley, *Report of a Conference on Strength of Solids* (The Physical Society, London, 1948), p. 57; also J. App. Phys. **18**, 1029 (1947).

[6] C. S. Barrett, Trans. A. I. M. E. **161**, 15 (1945).

[7] Smoluchowski, Lucht, and Mann, Phys. Rev. **70**, 318 (1940).

metal consisting of perfect individual crystals is illustrated in Fig. 1. A plane-parallel incident beam of wavelength λ is diffracted through an angle 2θ by crystals satisfying the Bragg condition (such as 1 and 3). If $2\theta > \alpha_{ob}$, crystals 1 and 3 will appear dark in the image with the intensity being determined by incoherent scattering only. Since an objective diaphragm of diameter 0.002 in. was employed in this work, it is evident that for all values of d less than about 9A, $2\theta > \alpha_{ob}$ for $\lambda = 0.054$A (50 kv electrons). Regions in the specimen which diffract strongly will thus appear dark in the electron image under the experimental conditions employed in this study. The relative intensities and the manner in which they are affected by thickness of the crystal and by orientation changes is a problem in the application of the dynamical theory.

I. DYNAMICAL THEORY OF ELECTRON DIFFRACTION

Since the interpretation of the electron images requires direct reference to the results of the dynamical theory, it seems advisable to present first the theory and then the experimental results.

The theory as developed here is concerned with finding the wave vectors and wave functions for an electron moving in the periodic potential inside the crystal.

The permissible crystal wave vectors are determined by:

(a) The energy-wave vector relation found by solving the Schrödinger equation inside the crystal.

(b) The boundary conditions which must be observed at the crystal-vacuum interface. (Appendix III).

For the case of a single diffracted wave, the case treated in detail, it will be seen that there are two possible wave

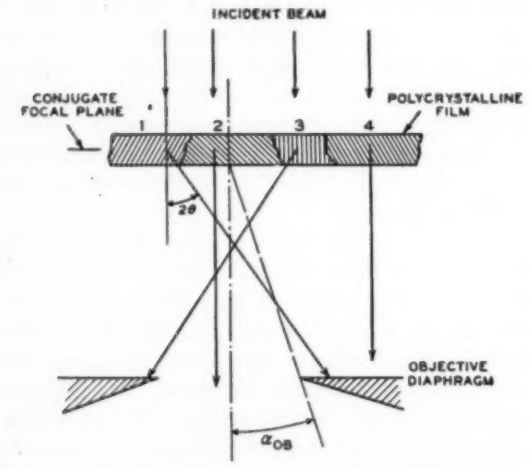


FIG. 1. Diffraction of electrons outside objective aperture by suitably oriented crystals in polycrystalline film. Crystals 1 and 3 would appear dark in final image.

vectors for the diffracted wave and two for the transmitted wave (see Fig. 5). There are accordingly two diffracted waves of slightly different wave-length giving rise to beats and resulting in a periodic dependence of intensity upon the thickness of the crystal.

Some of the mathematical treatment not essential to the continuity of the discussion will be found in appendices or references. It seems advisable, however, to first briefly review several relations in common usage in wave mechanics before considering the diffraction problem.

The time-free, non-relativistic Schrödinger equation in Cartesian coordinates is

$$\nabla^2\Psi + (8\pi^2 m/h^2)[E - V(x, y, z)]\Psi = 0, \qquad (1)$$

where Ψ is the wave function, E the total energy of the electron, $V(x, y, z)$ the potential energy, h is Plancks' constant and m the rest mass of the electron. Ψ is subject to the condition that it be everywhere single-valued and continuous and that $\int_\Omega \psi\psi^* \, d\Omega$ exist where $d\Omega$ is an element of volume and Ψ^* denotes the complex conjugate of Ψ.

For an electron in field-free space for which $V(x, y, z) = V_0$ (a constant), the general solution of (1) is a superposition of plane waves

$$\Psi^0 = \sum_j A_j e^{i(K_j \cdot r)}, \qquad (2)$$

where A_j is the amplitude, r the position vector and K_j the wave vector ($|K_j| = 2\pi/\lambda_j$ where λ denotes the wave-length). The energy and wave vector are related by the familiar expression

$$E^0 = h^2|K|^2/8\pi^2 m + V_0. \qquad (3)$$

The group velocity, U, for a de Broglie wave is given by

$$U = 2\pi(\partial\nu/\partial k) = (2\pi/h)[dE(k)/dk], \qquad (4a)$$

since $E = h\nu$ where ν is the frequency. For free electrons, (4a) becomes, using (3),

$$U^0 = hk/2\pi m = p/m, \qquad (4b)$$

where p is the classical momentum. U is thus the velocity of the electron. Equations (2)–(4) apply to electrons before entering and after leaving the crystal.

The quantity $\Psi\Psi^* d\Omega$ is interpreted as the probability of finding the electron in the element of volume $d\Omega$. $\Psi\Psi^*$ is the probability density or electron concentration and corresponds to the intensity that is observed experimentally.

A. Crystal Wave Functions

Inside the crystal, an electron is subject to the potential function $V(x, y, z)$ which is periodic in three dimensions with the period of the lattice. For an infinite, perfect crystal the potential distribution is assumed to be representable by a three-dimensional Fourier series

$$V(x, y, z) = -(V_0 + \sum_g{}' v_g e^{2\pi i(g \cdot r)}), \qquad (5)$$

where, for brevity, $\sum_g{}'$ indicates summation over the triple indices and the prime means $g_1 = g_2 = g_3 = 0$ is excluded (Appendix I). V_0 is the value of V when $g_1 = g_2 = g_3 = 0$ and is the mean or inner potential of the crystal.

Although the general solution of (1) is not known, it is necessary that for $V(x, y, z)$ periodic, Ψ must also be periodic and with the same period. It can be shown that such a solution of (1) is the Bloch function[8,9,10]

$$\varphi_K = e^{i(K\cdot r)}\mu_K(r), \tag{6}$$

where K is the wave vector, r the position vector and $\mu_K(r)$ is a periodic function with the period of the potential. The Bloch solution is simply an amplitude modulated plane wave. The assumption of the solution (6) does not uniquely determine the wave vector K and hence a Bloch function must be written for each permissible wave vector.

Since $\mu_K(r)$ is periodic with the period of the lattice it likewise is assumed representable as a Fourier series

$$\mu_K(r) = \sum_m C_{Km} e^{2\pi i(m\cdot r)}, \tag{7}$$

where m denotes a vector of the reciprocal lattice ($m = m_1b_1 + m_2b_2 + m_3b_3$) and the summation is over the triple indices.

The approximation of nearly free electrons,[10] or loose binding, is employed and is justified for the case of fast electrons for which $E \gg V$. This assumption requires that as the Fourier coefficients in (5) approach zero, the solution of (1) shall pass over to the plane wave solution (2) for free electrons. The Bloch function[6] satisfies this requirement.

The problem now consists in finding the Fourier amplitude coefficients C_{Km} in terms of the potential coefficients v_g.** Substituting (7), (6), and (5) into (1) and eliminating the exponentials using the orthogonality property yields the "Dispersion Equation"

$$(|K+2\pi s|^2 - (8\pi^2m/h^2)(E-V_0))C_{Ks} = (8\pi^2m/h^2)\textstyle\sum_g{}' v_g C_{K\lambda}, \tag{8}$$

where $s = s_1b_1 + s_2b_2 + s_3b_3$ is a vector of the reciprocal lattice and $\lambda = s - g$. An equation similar to (8) must be written for each s under consideration.

Equation (8) can also be written as

$$\beta_j C_j = \frac{8\pi^2 m}{h^2} \sum_{\substack{m \\ m \neq j}} v_{j-m} C_m, \tag{8a}$$

with

$$\beta_j \equiv \left[|K+2\pi j|^2 - \frac{8\pi^2 m}{h^2}(E - V_0) \right].$$

If only the amplitude coefficients C_0, C_g, C_s and C_λ*** are to be considered (8a) leads to the set of equations

$$\begin{bmatrix} (h^2/8\pi^2m)\beta_0 & -v_g^* & -v_s^* & -v_\lambda^* \\ -v_g & (h^2/8\pi^2m)\beta_g & -v_\lambda^* & -v_{\lambda-g}^* \\ -v_s & -v_\lambda & (h^2/8\pi^2m)\beta_s & -v_g \\ -v_\lambda & -v_{\lambda-g} & -v_g^* & (h^2/8\pi^2m)\beta_\lambda \end{bmatrix} \begin{Bmatrix} C_{K0} \\ C_{Kg} \\ C_{Ks} \\ C_{K\lambda} \end{Bmatrix} = 0. \tag{9}$$

Equation (9) has been written as if s and g were double indices since a three-dimensional matrix would be required to correctly represent the system. The presence of the amplitude $C_{K\lambda}$ in (9) is to be noted even if $v_\lambda = 0$. This point will be discussed later. The array (9) is the energy matrix of the system the secular equation for which determines the eigenvalues or proper frequencies.

The case where only one amplitude coefficient, say C_{K0}, in Eq. (8), is important is of little interest here since this is equivalent to neglecting the diffracted waves. The case of two strong amplitudes C_{K0} and C_{Kg}, is of primary interest however, and will be considered in detail. Equation (9) now reduces to

$$\begin{pmatrix} (h^2/8\pi^2m)\beta_0 & -v_g^* \\ -v_g & (h^2/8\pi^2m)\beta_g \end{pmatrix} \begin{Bmatrix} C_{K0} \\ C_{Kg} \end{Bmatrix} = 0, \tag{10}$$

and the secular equation yields the roots

$$E = V_0 + \frac{h^2}{16\pi^2m}(|K|^2 + |K_g|^2) \pm \left(|v_g|^2 + \frac{h^4}{16\pi^2m^2}[(K\cdot g) + \pi|g|^2]^2 \right)^{\frac{1}{2}}, \tag{11}$$

which is the desired relation between the energy E and wave vector K. In (11), $Kg = K + 2\pi g$ so that $|Kg|^2 - |K|^2 = 4\pi[(K\cdot g) + \pi|g|^2]$. The expression for the energy (Eq. (11)) is plotted for the case of one dimension in Fig. 2. The negative radical in (11) corresponds to the region near A and the positive radical to the region near B. Equation (11) applies only in the neighborhood of the discontinuity occurring at $|K|^2 = |K_g|^2$.**** When $|K|^2 = |K_g|^2$, (11) becomes $E = E^0 \pm |vg|$ (E^0 from (3)) or the system is degenerate. The treatment of diffraction phenomena amounts to the application of perturbation tkeory to a degenerate system.

Figure 2 is the familiar E *vs.* K curve encountered in the band theory of solids; i.e., the energy range $O < E < A$ is allowed, $A < E < B$ is forbidden and $E > B$ allowed for the valence electrons in the crystal, etc. The population of the levels in the allowed zones is of no interest here but the energy gap $\Delta E = 2|v_g|$ is important.

Equation (10) immediately gives for the amplitude coefficient C_{Kg};

$$C_{Kg} = \alpha_g C_{K0} \quad \text{where} \quad \alpha_g = (8\pi^2m/h^2)(v_g/\beta_g) \tag{12}$$

and the Bloch functions are of the form,

$$\varphi_K = C_{K0}e^{i(K\cdot r)} + C_{Kg}e^{i(K_g\cdot r)}. \tag{13}$$

[8] H. Fröhlich, *Elecktronentheorie der Metalle* (Verlag. Julius Springer, Berlin, 1936).
[9] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, London, 1936).
[10] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Clarendon Press, Oxford, 1936).

** See references (8)–(10).
*** $\lambda = s - g$.
**** (11) reduces to (3) for $(h^4/16\pi^2m^2)[(K\cdot g) + \pi|g|^2]^2 \gg |v_g|^2$.

It remains now to ascertain which vectors K are possible in the crystal. Since the electron beam is of fixed energy E and since inelastic scattering is neglected, it is evident first of all that only those values of K consistent with Eq. (11) are permissible. Let the wave vector associated with the negative sign in (11) be K and that associated with the positive sign be K'. K and K' are the only two crystal wave vectors possible when the energy and incident wave vector are fixed. It is reasonable to suppose that K and K' differ by another vector ΔK or

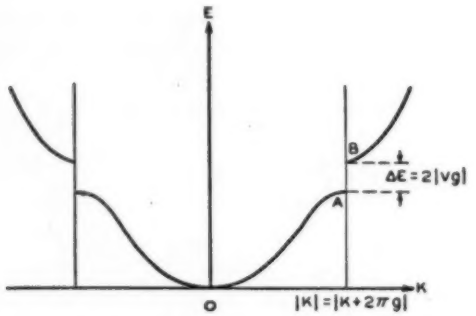


FIG. 2. Plot of energy, E, vs. wave number $|K|$, along K vector showing discontinuity when $|K|=|K+2\pi g|$ or when Bragg condition is realized.

that

$$K'=K+\Delta K. \tag{14}$$

The value of ΔK is to be determined from (11) where K' corresponds to the positive sign and K to the negative sign thus giving two expressions which are equal. As will be seen, ΔK is a small quantity and $|\Delta K|^2$ can be neglected. Also, if the beam is not too obliquely inclined to the surface, $(\Delta K\cdot g)$ can be neglected.† The result is then

$$\Delta K\cdot K=-\{[(8\pi^2m/h^2)v_0]^2 +4\pi^2[(K\cdot g)+\pi|g|^2]^2\}^{\frac{1}{2}}. \tag{15}$$

In Bethe's theory, ΔK is called the "anpassung" and is simply the difference between the two crystal wave vectors consistent with a fixed total energy; i.e., the beat wave vector. ΔK must satisfy still another condition, as will be seen, if K and K' are to satisfy the boundary requirements.

The amplitude coefficients $C_{K'0}$ and $C_{K'g}$ are related in the same way as were C_{K0} and C_{Kg} or

$$C_{K'g}=\alpha_g'C_{K'0} \quad \text{with} \quad \alpha_g'=(8\pi^2m/h^2)(v_0/\beta_g'). \tag{16}$$

Using (11), the values of β_g and β_g' are

$$\beta_g=\tfrac{1}{2}(|K_g|^2-|K|^2) -\{[(8\pi^2m/h^2)v_0]^2+4\pi^2[(K\cdot g)+\pi|g|^2]^2\}^{\frac{1}{2}}, \tag{17a}$$

$$\beta_g'=\tfrac{1}{2}(|K_g'|^2-|K'|^2) +\{[(8\pi^2m/h^2)v_0]^2+4\pi^2[(K'\cdot g)+\pi|g|^2]^2\}^{\frac{1}{2}}. \tag{17b}$$

† See Fig. 5.

The normalized Bloch functions sought are thus

$$\varphi_K=\frac{1}{C_{K0}[(1+\alpha_g^2)\Omega]^{\frac{1}{2}}}[e^{i(K\cdot r)}+\alpha_g e^{i(K_g\cdot r)}], \tag{18a}$$

$$\varphi_{K'}=\frac{1}{C_{K'0}[(1+\alpha_g'^2)\Omega]^{\frac{1}{2}}}\{e^{i(K'\cdot r)}+\alpha_g' e^{i(K_g'\cdot r)}\}. \tag{18b}$$

Higher order approximations can be found by placing the values of C_{Kg} and $C_{K'g}$ back into (8) (using (17a) and (17b)) and solving again for the energy. The correction is not large, however, and does not add materially to the theory for two strong amplitudes. The case of three strong amplitudes is of far greater importance.

The crystal wave function; then, for the case of two strong amplitudes is a linear combination of (18a) and (18b).

$$\psi_c=a\varphi_K+b\varphi_{K'}, \tag{19}$$

where a and b are to be determined by the boundary conditions.

B. Diffraction Conditions

Before evaluating the arbitrary constants in Eq. (19), it is instructive to consider the diffraction conditions and their connection with Fig. 2 in some detail. The wave vectors employed in the preceding discussion are vectors in reciprocal space and hence can be directly employed in the Ewald construction for the diffraction conditions. Figure 3 shows a section of the reciprocal lattice with the wave vector K drawn from the origin, O, and oppositely directed to the incident wave. K_g represents the diffracted wave vector. The Laue condition in reciprocal space is simply

$$|K|^2=|K_g|^2, \tag{20}$$

or

$$(K\cdot g)+\pi|g|^2=0=(K_g\cdot g)-\pi|g|^2,$$

which is equivalent to the Bragg law.

If the diffraction condition is not exactly satisfied, $|K|^2\neq|Kg|^2$, a simple geometric construction shows that

$$2\pi/|K|^2[(K\cdot g)+\pi|g|^2]=(\Delta\theta)\sin 2\theta, \tag{21}$$

where θ_0 is the Bragg angle and $\Delta\theta$ is the deviation from the Bragg angle.

It is evident now that the discontinuity in the E vs. K relation (Fig. 2) occurs when the wave vectors K and K_g fulfill the Laue condition (20). The second term under the radical of (11) is simply the deviation from the Bragg angle.

Returning to Fig. 3, it is seen that the diameter, LM, of the Ewald sphere is the perpendicular bisector of the reciprocal lattice vector, g. The incident wave vector, K, can be translated to $O'R$; an equivalent construction. In three dimensions LM will be a plane and is the Brillouin[11]

[11] L. Brillouin, *Wave Propagation in Periodic Structures* (McGraw-Hill Book Company, Inc., New York, 1946).

zone boundary associated with the reciprocal lattice vector, g. *The Laue condition (20) now can be stated as requiring that the wave vector K terminate on the Brillouin zone boundary associated with the reciprocal lattice vector g.* If the perpendicular bisectors of the reciprocal lattice

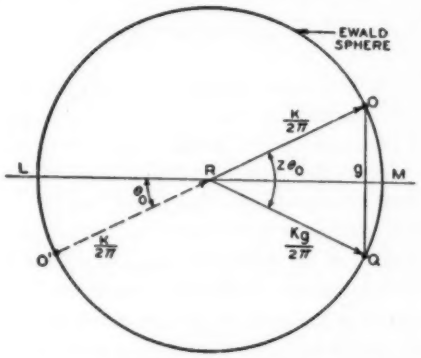


FIG. 3. Ewald construction for diffraction condition in reciprocal space. G is a reciprocal lattice point and O is the origin. $K/2\pi$ denotes incident wave vector and $K_g/2\pi$ is diffracted wave vector. (The Q on the right hand side of the figure should be G.)

vectors are drawn to form the smallest polyhedron centered at the origin, the volume so formed is the first Brillouin zone. The volume between the first zone and the second largest polyhedron is the second zone, etc. The first three zones for a simple square lattice are shown in Fig. 4. In passing from the first zone to the second, a boundary must be crossed presenting an energy discontinuity $\Delta E=2|v_g|$ (Fig. 2). If $\Delta E=0$, there is no zone boundary so that the construction of Brillouin zones requires a knowledge of the structure factor of the lattice. Zones for several crystals will be found in reference 10.

The assumption of only two strong amplitudes leading to (10) is equivalent to assuming that the Ewald sphere passes through one and only one reciprocal lattice point. It can be seen in Fig. 3 that if the Ewald sphere passes through another reciprocal lattice point (different from G) R will be the junction of two Brillouin zone boundaries. That is, if the incident wave vector K terminates near a Brillouin zone corner, then two or more diffracted waves must be considered. Equation (9) applies at the corner produced by the Brillouin zone boundaries associated with v_g, v_s, and v_λ. The secular equation from (9) leads to a quartic in the energy yielding four roots and three distinct energy gaps (Appendix II). In a zone corner, the energy gaps are no longer given by twice the Fourier coefficients due to perturbation of the levels. Although v_λ may be explicitly set equal to zero in (9), there is an effective gap and a non-zero amplitude coefficient $C_{K\lambda}$. The "perturbation amplitude" $C_{K\lambda}$ amounts to a "forbidden" reflection on the basis of the classical structure factor. It appears now that the fundamental quantity determining the existence of a diffracted wave is the energy gap ΔE and only in the simple case does $\Delta E=2|v_g|$.

It may well be that the "forbidden" (222) reflection in silicon and germanium is an example of the behavior near a zone corner. If $s\equiv(331)$ and $g\equiv(11\bar{1})$ then $\lambda=(222)$ and there will be an effective energy gap $\Delta E_{(222)}$ and a non-zero amplitude $C_{(222)}$. The analysis has not been carried sufficiently far for a comparison between theory and experiment as yet.

Whether or not the incident wave vector terminates near a Brillouin zone boundary or near a corner will depend upon the direction and magnitude of K and upon the details of the zone structure. The reflections which can occur can be determined by drawing a sphere of radius $|K/2\pi|$ centered at the origin of reciprocal space. Wherever this sphere cuts a zone boundary or corner, a diffracted wave vector must be constructed. If the sphere cuts through a corner, classically forbidden reflections may occur and the relative intensities of all the reflections concerned may be modified.

It remains to be seen how well the theory applied to zone corners will agree with experiment. The calculation of the energy levels at Brillouin zone corners and the construction of the corresponding wave functions is somewhat tedious but is of major importance in electron diffraction theory.

C. Boundary Conditions

Before the intensities of diffraction can be calculated, the coefficients in Eq. (19) must be evaluated. Also, the wave vectors K and K' must conform to the boundary requirements which serve to place further restrictions on them in addition to the fixed energy. The boundary condition requires that at the crystal surface the tangential components of all wave vectors (both crystal and

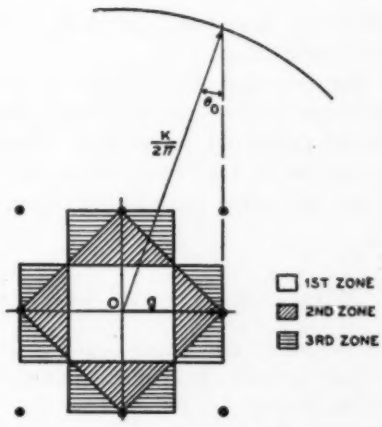


FIG. 4. First three Brillouin zones for a simple square lattice. In actual case, $|g|$ is of the order of $0.5A^{-1}$ and $|K/2\pi|$ about $19A^{-1}$ for 50 kev electrons.

vacuum) must be equal or differ only by the tangential component of the reciprocal lattice vector. (Appendix III). Using this condition, the crystal wave vectors K and K' can be drawn as shown in Fig. 5. The plane of the figure is the plane containing the surface normal and

incident wave vector. The incident wave vector in the crystal, K, is drawn in nearly the same direction as the incident vacuum wave vector K_0 and terminates near a Brillouin zone boundary but not near a corner. K and K_0 would be parallel but for refraction due to the inner potential, V_0.

Since $|\Delta K/2\pi|$ is in general much smaller than $|g|$, it follows that ΔK must be normal to the surface as drawn in Fig. 5. It is assumed that no electrons enter the vacuum from the front surface (Appendix III) and hence

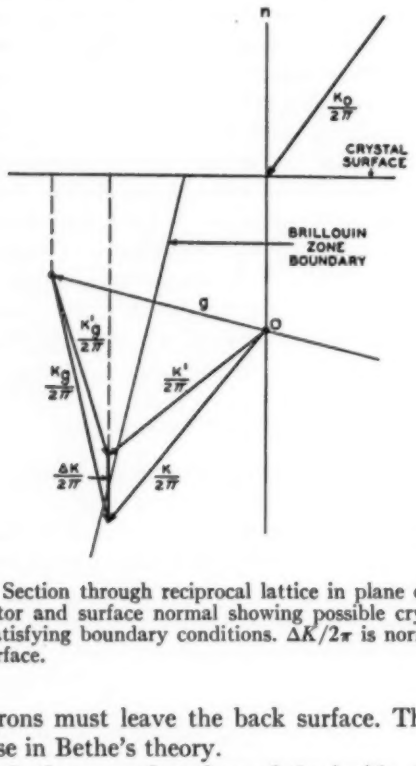


FIG. 5. Section through reciprocal lattice in plane of incident wave vector and surface normal showing possible crystal wave vectors satisfying boundary conditions. $\Delta K/2\pi$ is normal to the crystal surface.

all electrons must leave the back surface. This is the Laue case in Bethe's theory.

At $z=0$, the wave functions of the incident vacuum and incident crystal waves must be equal and the wave function of the diffracted waves zero. Applying these conditions (Appendix III) finally gives for the wave function of the diffracted wave in the crystal

$$\varphi_g=(\alpha_g\alpha_g'/\alpha_g-\alpha_g') \times\{\exp[i(K_g\cdot r)]-\exp[+i(K_g'\cdot r)]\}. \quad (22)$$

The same boundary conditions apply at the back or exit surface although all the coefficients are determined at the incident surface.

D. Diffracted Intensities

The intensity of the diffracted wave can be computed at once from Eq. (22) to give

$$\varphi_g\varphi_g^*\simeq|A|^2(\alpha_g\alpha_g')^2\sin^2\tfrac{1}{2}(\Delta K\cdot r), \quad (23)$$

since $(\alpha_g-\alpha_g')\simeq 2$. For nearly normal incidence, ΔK and r are practically parallel and so using (15) and (21) and the values of α_g and α_g' the final expression is

$$\frac{I_g}{I_0}=\frac{\varphi_g\varphi_g^*}{|A|^2}\simeq\left[\frac{V_{hkl}}{E}\right]^2 \times\left\{\frac{\sin\tfrac{1}{2}|K|z[(V_{hkl}/E)^2+(\Delta\theta)^2\sin^2 2\theta_0]^{\frac{1}{2}}}{[(V_{hkl}/E)^2+(\Delta\theta)^2\sin^2 2\theta_0]^{\frac{1}{2}}}\right\}^2, \quad (24)$$

which is just the form given by MacGillavry.[2]†† In review, (24) represents the relative intensity of the diffracted beam when the incident beam is nearly normal to the surface of the crystal and is valid for fast electrons in a perfect crystal when the wave vector terminates near a Brillouin zone boundary but not near a corner. Attenuation due to inelastic scattering is neglected. Even so, comparison between theory and experiment can be made semi-quantitatively since several of the dynamical effects are not obscured by attenuation.

The two outstanding phenomena predicted by the theory are the periodic variation of the diffracted intensity with distance in the crystal and with deviation from the Bragg angle. The first is most easily seen when $\Delta\theta=0$ so that (24) becomes

$$I_2/I_0=\sin^2\pi z/z_0 \quad \text{where} \quad z_0=(\lambda E/V_{hkl}). \quad (25)$$

The diffracted intensity, I_g, and the transmitted intensity $I_T=I_0-I_g$ are plotted as a function of z in Fig. 6. The intensities actually observed are determined by the location of the back or exit face of the crystal. If the crystal were cut at $z=D$, the intensities of the exit vacuum waves would be determined by the value of I_g at $z=D$. It will be noted that the period of the intensity variation, $z_0=\lambda E/V_{hkl}$, is of the order of two or three hundred Angstroms in this work. z_0 has been checked experimentally with electron images of small regular crystals and found to be approximately correct.[12–14] In judging the experimental results, the degree to which the assumptions of the theory are realized must be kept in mind.

In general, relative maxima in the diffracted intensity

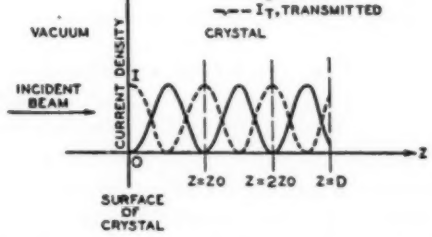


FIG. 6. Periodic variation of reflected intensity with distance of propagation through crystal.

†† (24) is also of the form found in the dynamical theory of x-ray diffraction. Cf. W. Zachariasen, *Theory of X-ray Diffraction in Crystals*, Sec. III, 8, 9, 10, 11. (John Wiley and Sons, Inc., New York, 1945).

[12] E. Kinder, Naturwiss. **31**, 149 (1943).
[13] R. D. Heidenreich and L. Sturkey, J. App. Phys. **16**, 97 (1945).
[14] C. E. Hall, J. App. Phys. **19**, 198 (1948).

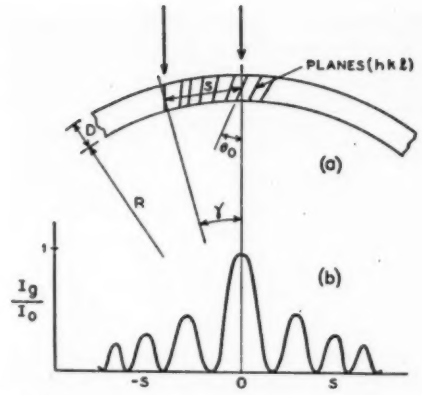


FIG. 7. (a) Geometric construction for an ideally bent crystal. (b) Primary and secondary diffraction maxima predicted for the bent crystal.

will occur whenever

$$(D/\lambda)[(V_{hkl}/E)^2+(\Delta\theta)^2\sin^2 2\theta_0]^{\frac{1}{2}} = n/2;\ n=1, 3, 5, 7\cdots, \quad (26)$$

by Eq. (24) where D is the thickness of the crystal. In the case of electron images of thin crystal sheets, Eq. (26) can be said to determine the "extinction contours" which will be exhibited in Part II. Equation (26) can be written

$$(\Delta\theta)_n=\pm\left[\left(\frac{d_{hkl}}{D}\frac{n}{2}\right)^2-\left(\frac{d_{hkl}}{\lambda}\frac{V_{hkl}}{E}\right)^2\right]^{\frac{1}{2}}, \quad (26a)$$

giving the values of $\Delta\theta$ at which relative maxima will occur. For large n, the angular spacing of adjacent maxima is

$$(\Delta\theta)_{n+2}-(\Delta\theta)_n\simeq(d_{hkl}/D). \quad (26b)$$

Equation (26b) has been semi-quantitatively verified.[2,15] The maxima predicted by (26a) sometimes are observed in conventional electron diffraction patterns as a series of fringes on either side of a Kikuchi line and have been demonstrated both by transmission[16] through thin crystal flakes and by reflection.[5]

It is perhaps worth while at this point to say a few words concerning the physical aspects of the dynamical theory. When an incident electron enters a crystal such that the Bragg condition is nearly satisfied, the incident wave vector and energy correspond to a forbidden energy range. In the forbidden range, the wave vectors are actually imaginary. This situation results in an interaction with the crystal to produce two real wave vectors, K and K', corresponding to the same total energy. The kinetic energy associated with these two wave vectors is not the same but differs by an amount $2|v_g|$. There are thus two waves of slightly different wave-length traveling almost in the same direction which interfere and give rise to beats. The beat wave-length, λ_b, is given by $\lambda_b=2\pi/|\Delta K|=\lambda E/V_{hkl}$ when $\Delta\theta=0$. Both the transmitted and diffracted crystal waves exhibit this beat wave-length but with $\pi/2$ phase difference. When the thickness of the crystal is an integral multiple of λ_b, no energy is propagated in the diffracted wave as seen in Fig. 6. In a sense, the behavior can be likened to the transmission of light through thin films of the order of the wave-length in thickness. A wedge shaped crystal exhibits the electron analog of Newton's rings.

The two wave vectors K and K' are the electron analog of double refraction in light optics. Since the refractive index, μ, is given by $\mu=|K/K_0|$ where K_0 is the incident, vacuum wave vector, the result is

$$\mu=[1+(V_0\pm V_{hkl}/E)]^{\frac{1}{2}}.$$

This expression has been advanced by Sturkey††† to explain the multiple Bragg spots observed in the diffraction patterns of M_gO smoke. The interaction between the ordinary and extraordinary rays must be taken into account, however.

In the case of Debye-Scherrer or ring patterns produced by an assembly of randomly oriented crystallites, Eq. (24) must be integrated over all orientations. Blackman†††† has carried out the integration and shows that the intensities observed are in fair agreement with experiment.

E. Application to Electron Images

The application of the dynamical theory to the electron images to be presented in Part II is primarily con-

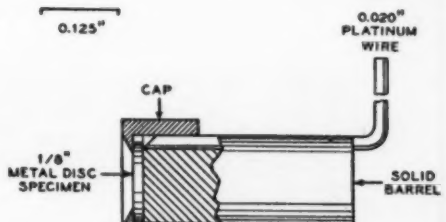


FIG. 8. "Teflon" electro-thinning specimen holder. Cap to be slide fit on barrel. Internal diameter of cap exceeds that of specimen disk by about 0.015".

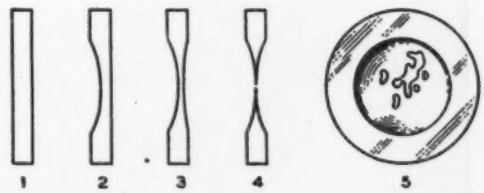


FIG. 9. Stages in electro-thinning of metal disk. (1) Profile of initial 0.005" thick disk. (2) First surface electropolished in holder. (3) Disk reversed in holder and second surface polished. (4) Stage at which thinning is halted. (5) Front view of disk at stage 4 showing holes.

[15] N. Davidson and J. Hillier, J. App. Phys. **18**, 499 (1947).
[16] W. Kossel and G. Möllenstedt, Ann. d. Physik **36**, 113 (1939). See also: I. Ackerman, Ann. d. Physik **2**, 19 (1948).

††† L. Sturkey, Phys. Rev. **73**, 183(L) (1948).
†††† M. Blackman, Proc. Roy. Soc. **A173**, 68 (1939).

cerned with Eqs. (26). If a slab of perfect crystal of uniform thickness D is imaged in the electron microscope, the image intensity should be uniform over the irradiated area of the crystal if the incident beam is plane-parallel. The intensity in the image should then vary with relative orientation of crystal and beam according to Eq. (24). However, if the crystal is bent or varies in thickness, then the intensity distribution in the image will not be uniform. Regions in which Eq. (26) is satisfied will appear dark in the image (Fig. 1).

If an ideal crystal slab of thickness D is bent on a radius R and if it is assumed that no local distortions occur, than a first approximation to the image is obtained by assuming a continuously varying orientation. The geometric construction is shown in Fig. 7(a). The location of the relative maxima, s, measured from the line $\Delta\theta=0$, will be given by $s=\gamma(R+D/2)$ which, in conjunction with (32), yields (since $\gamma=\Delta\theta$)

$$s=\pm(R+D/2)\left[\left(\frac{d_{hkl}}{D}\frac{n}{2}\right)^2-\left(\frac{d}{\lambda}\frac{V_{hkl}}{E}\right)^2;n\right]^{\frac{1}{2}}=1,3,5,7\cdots. \quad (27)$$

A plot of the intensity as a function of distance from the primary line is sketched in Fig. 7(b). On this basis, the

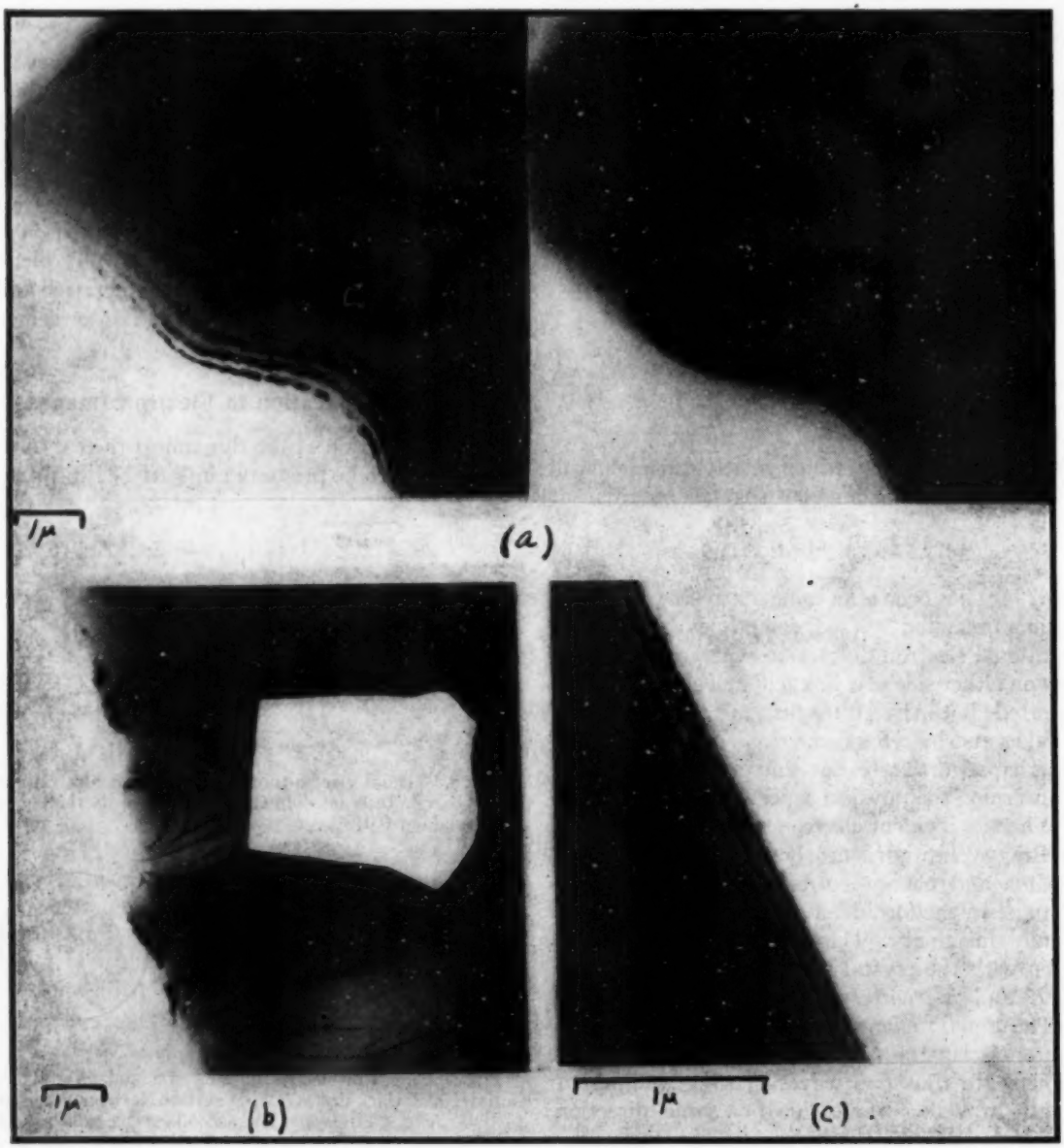


FIG. 10. Extinction contours produced by wedge sections near edges of holes. The dark fringes are the regions in which diffraction occurred (see Fig. 1). (a) Stereoscopic pair showing effect of orientation of section relative to electron beam. (b) Non-uniform edge. (c) Relatively uniform edge with large thickness gradient.

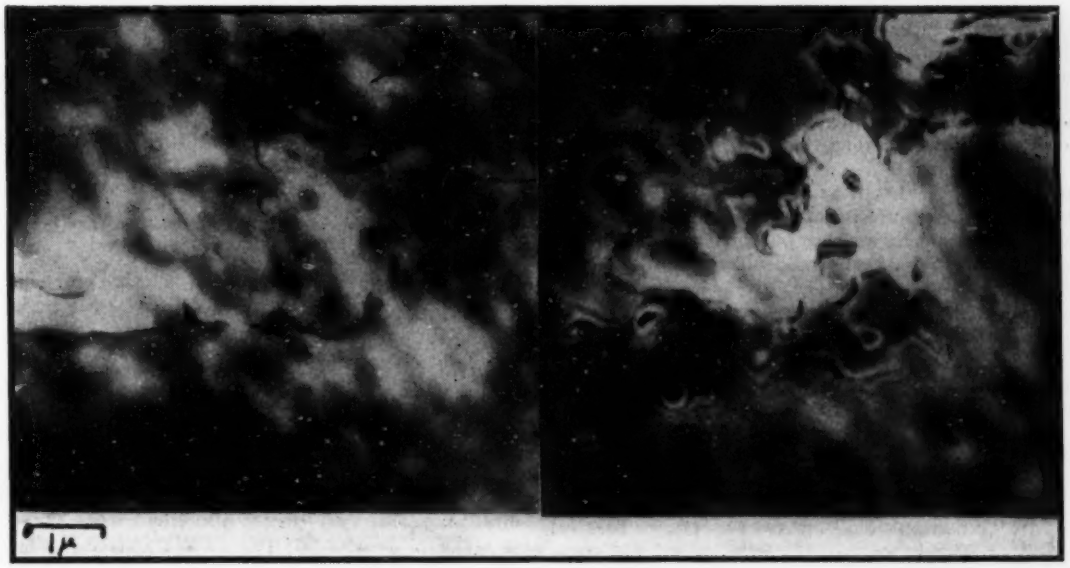


FIG. 11. Stereomicrographs of a pure aluminum section with local variations in thickness resulting from improper polishing conditions. The thickness gradients are evident through both incoherent scattering and extinction contours.

electron images for cylindrically bent and spherically bent crystals would exhibit parallel fringes and concentric, circular fringes respectively. Some idea, then, as to whether or not the diffraction maxima observed are the result of simple, ideal bending of the crystal and thickness variations or whether local distortions must be assumed can be gained from studying the extinction contours. This is one of the problems in interpretation to be met in Part II.

The present experimental results show that a detailed analysis of the extinction contours will be quite complex due to irregular bending and local distortions accompanying plastic deformation. The theoretical treatment given is merely a starting point in the attempt to interpret the images.

II. EXPERIMENTAL METHODS AND RESULTS

A. Preparation of Sections

The electro-thinning technique was first developed for high purity aluminum and then extended without modification to an aluminum-copper alloy. The general procedure will likely be the same for any metal or alloy but with the electrolyte and polishing conditions varied

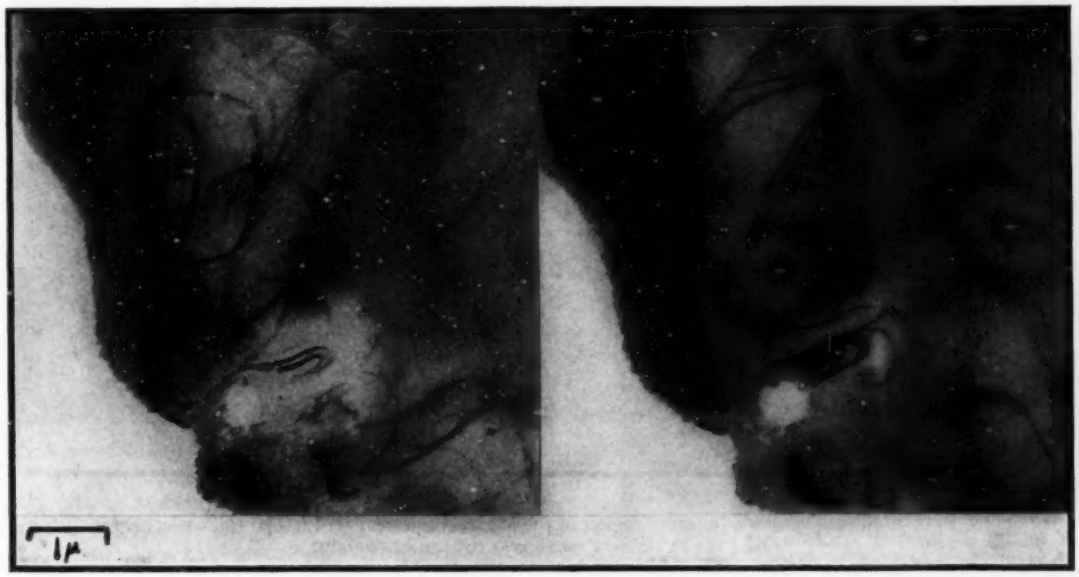


FIG. 12. Extinction contours in flash annealed section of pure aluminum. The contours end abruptly at what is presumably a grain boundary.

for the particular alloy under consideration. The pure aluminum and Al-4 percent Cu sections were prepared as follows:

(1) The metal is first rolled into sheet about 0.005-in. thick.
(2) Recrystallization, annealing, solution heat treatment, etc., are carried out.
(3) A piece of the sheet about 1 in. square is lightly electropolished to remove surface inclusions and oxide.
(4) Disks $\frac{1}{4}$ in. in diameter are then punched from the sheet.
(5) The electro-thinning is next carried out with the disk held in the holder shown in Fig. 8. The material from which the holder is made, polytetrafluorethylene,* was chosen because of its resistance to strong acids and organic solvents. The disk is held against the end of the barrel by a cap which exposes only the central region of the disk to the electrolyte. A piece of 0.020-in. platinum wire runs through a groove in the barrel and makes electrical contact with the disk by means of a sharp point.

The electrolyte (for aluminum) is a modification of that given by Waisman[17] and consists of 1 part methyl alcohol, 1 part conc. HNO_3 and 1 cc of conc. HCl to 50 cc of the mixture. The electrolyte polishes at a lower voltage with less gas evolution than does the formula given by Waisman. The HCl reduces the tendency to form oxide.

Polishing is carried out in a 100 cc bath of electrolyte in a glass dish with a pure aluminum cathode. A direct current source is employed with the usual polishing conditions being a 5-volt drop between electrodes and a current of about 1 ampere. The specimen holder is held by a clip on the extremity of the platinum wire and is maintained with its long axis horizontal. The face of the disk is usually about $\frac{1}{2}$ in. from the cathode and near its edge so that the light from a source such as a spot light reflects from the disk. The holder is turned until the surface of the disk reflects brightly and can be observed during polishing with a 4× magnifier.

The process is begun by inserting the disk in the holder and thinning to about half-thickness at the center. The holder is then rapidly removed from the electrolyte and plunged in a stream of running water. It is then rinsed quickly in methyl alcohol and the cap removed and the disk shaken into 1:1 methanol-acetone. It is picked up from here on a large piece of screen and blotted on filter paper. The polished surface should appear very bright and clean when the procedure is carried out rapidly. The profile of the disk should now be as shown in stage 2 of Fig. 9. This profile is likely a result of the decrease in current density near the plastic due to the static charge accumulated on the surface of the cap. The disk is now replaced in the cap with the unpolished face out and polishing resumed. The surface must be carefully observed through a 4× glass so that the first puncture of the disk is seen and the holder im-

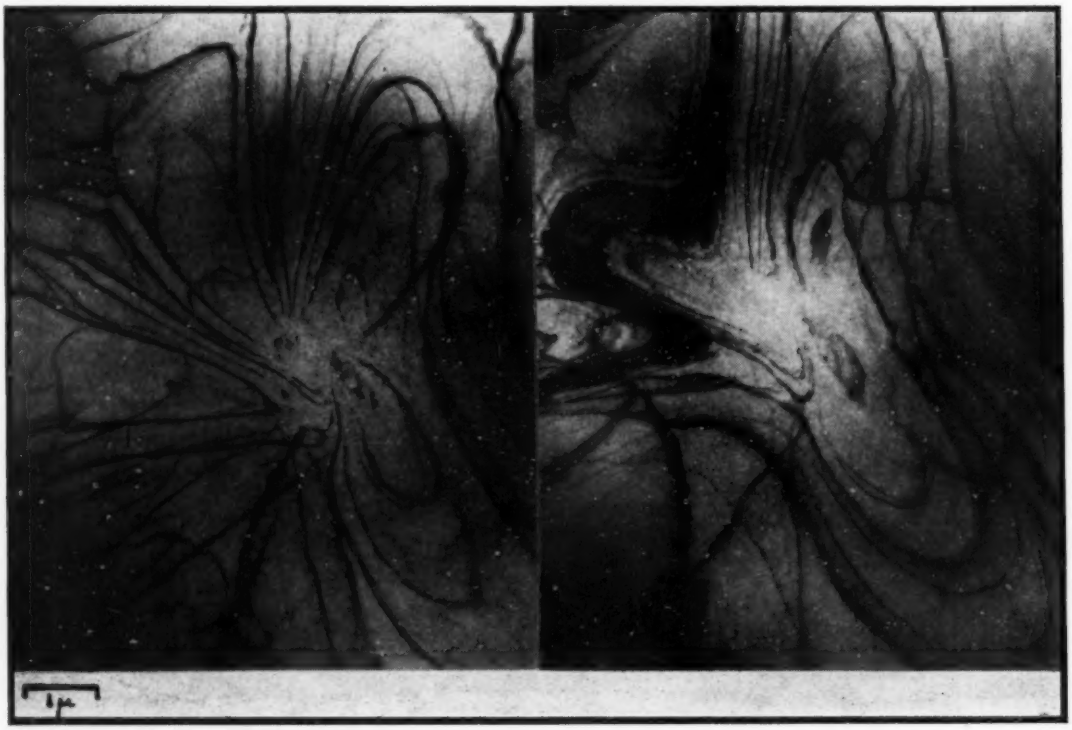


FIG. 13. Extinction contours in a different region of the same section as Fig. 12. The thickness here is such that the subsidiary maxima occur quite strongly.

* Sold by duPont under the trade name "Teflon." It is a tough, "waxy" material with a gray-blue color. Gold and platinum were first tried but proved unsatisfactory because of the low current efficiencies obtained.

[17] J. L. Waisman, Metal Progress, p. 607 (April, 1947).

FIG. 14. Region of Fig. 13 enlarged to show subsidiary maxima and discontinuities in the primary maxima. (See Eq. (27) and Fig. 7.)

mediately removed and plunged in water. The more promptly the action is halted upon the first appearance of a hole, the more thin area will be available. The same rinse and dry procedure is used as for the first face. The profile is similar to stage 4 of Fig. 9 while 5 shows the appearance of the disk face. There will generally be several holes visible with a light microscope at 50×. Handling of the thinned disk is relatively easy due to the thick rim. No mounting screens are required when the disk is placed in the specimen holder of the electron microscope.

The region around the holes will be found to be sufficiently thin to examine in the electron microscope at about 4000×, but with a progressive increase in thickness away from the edges of the holes. The usable sections are thus actually wedge-shaped and only a small percentage of the total area is usable. Good sections are not always obtained and it is often necessary to prepare 4 or 5 disks in order to produce a good one. The chief difficulties are usually among the following:

(a) Practically no thin regions, indicating that the polishing process was not halted soon enough.
(b) A general roughness of the surfaces giving rise to thickness variations. This is due to the polishing conditions not being right. Experimenting with conditions will enable the correct ones to be discovered.
(c) Hydrated oxide layers. These layers are a result of reaction of the acid with the metal during rinsing and can be minimized by rinsing and removing the disk from the cap as rapidly as possible. The disks should appear smooth and bright to the eye without indication of reaction products on the surface. In the electron microscope, the oxide layers will be seen to protrude into the holes, beyond the metal, and are usually quite transparent and sometimes granular in appearance.
(d) Distortions of the thin regions by gas evolution. Reduction of gas evolution to a minimum is necessary in preparing usable sections, particularly where the effects of plastic deformations are of interest.

A major disadvantage of the present technique is the relatively small area of usable section in specimens prepared by the method outlined. It appears that a sensitive method of determining the point at which thinning should be halted is required if reasonably large areas are

FIG. 15. Shadow electron microscope image of annealed aluminum section showing transmitted and diffracted images. The dark contours in the transmitted image can be paired with the white contours in the diffracted image.

to be produced. It is evident that once a puncture has occurred in the section, the thin regions about the hole are very rapidly consumed. A means for stopping the polishing process at the right point can likely be developed.

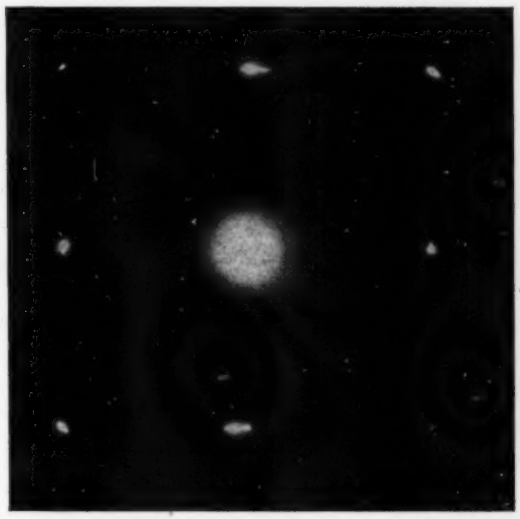


FIG. 16. Conventional electron diffraction pattern from a section similar to Fig. 15.

B. Effect of Thickness Variations

As predicted by the theory, local variations in thickness of the sections produce diffraction maxima. Figure 10 illustrates the appearance of the images for several thickness gradients. Figure 10(a) is a stereoscopic pair in one of which a set of planes was oriented so as to give rise to Bragg reflections while in the other only incoherent scattering occurred. The spacing and width of the maxima is determined by the thickness gradient. The broader the maxima and the more widely spaced they are, the smaller is the thickness gradient. This can be seen in Figs. 10(b) and 10(c). The thickness gradient in Fig. 10(b) is not uniform along the edge but, rather, is quite variable. The broad fringes suggest gullies or grooves running in from the edge with a small thickness gradient. Figure 10(c) shows a relatively uniform edge with a high thickness gradient resulting in closely spaced, sharp maxima (the spacing of the maxima is about 600A).

The effect of local thickness variations is illustrated in Fig. 11. The polish conditions were not correct in this instance and produced surfaces that were uneven thus giving variations in thickness. The thickness gradients can be seen by virtue of the incoherent scattering in one member of the stereoscopic pair of Fig. 11 and by means of the extinction contours in the other where the Bragg condition was fulfilled. Sections such as this are generally undesirable since other detail of interest is likely to be masked by the thickness effects.

C. Effect of Orientation Changes

The first thin sections prepared were from pure aluminum sheet which was flash-annealed in a Bunsen flame before thinning. Partial recrystallization occurred but the time was not long enough to produce a well annealed specimen.

The type of results obtained from the annealed sections are very similar to those previously found for mica[13] and are illustrated in Figs. 12 and 13. These are stereomicrographs used again to illustrate the effect of change of angle on the appearance of the image. Figure 12 is a simpler pattern then Fig. 13 and shows the primary extinction contours ($\Delta\theta=0$). The thickness of the section is reasonably uniform[b] and it is concluded that the extinction contours in this instance are due to bending of the crystal. A line at which the contours end abruptly is evident in Fig. 12 and is likely a grain boundary.

A much more complex pattern is illustrated in Fig. 13. A portion of the large number of contours can be ascribed to the subsidiary maxima predicted in Fig. 7 and shown more clearly in the enlarged region in Fig. 14. A rather peculiar distortion of the crystal, such as a dimple in the form of a clover leaf, is required to produce the symmetry of the pattern of Fig. 13.

It becomes evident that in order to start an analysis of the patterns it is essential to index the contours. This cannot be done from these images alone but requires a knowledge of the angles made by the diffracted beams

(a) (b)

FIG. 17. Electron micrograph and diffraction pattern from section of pure aluminum cold-worked and partially annealed showing domain structure.

[b] Except in the area of the small hole.

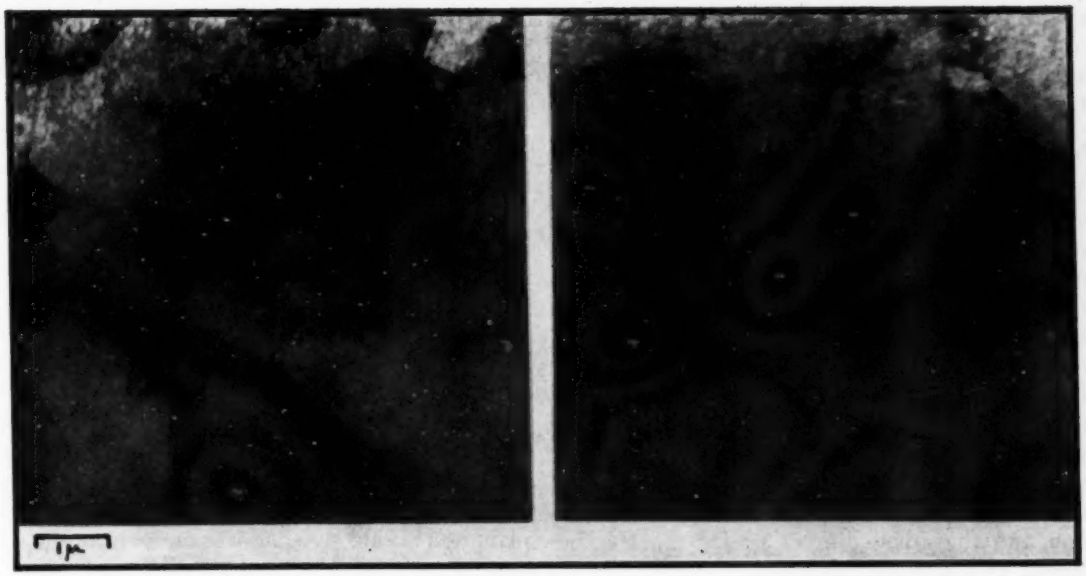


FIG. 18. Stereomicrographs of cold-worked aluminum illustrating disorientation of the domains. Some indication of the fine granular structure can be seen. The absence of extinction contours (such as Fig. 14) is to be noted.

However, some idea as to the radii of curvature involved in the bending can be obtained from Eq. (33). The separation of the contours indicated in Fig. 14 is about 670A. Taking V_{hkl} to be the order of 6 volts, $d\sim 2$A, $n=1$, $\lambda=0.054$A and $D\sim 300$A, Eq. (33) yields $R\sim 70\mu$ which demonstrates the order of magnitude.

In order to index the contours it is necessary to record the diffracted beams. The only method of accomplishing this at the present time is with the shadow microscope described by Hillier.[15] Dr. Hillier kindly offered to take some shadow images of thin sections in which the transmitted and diffracted images would be recorded. Figure 15 is representative of the type of result obtained. From the distance between the dark contour in the transmitted image and the bright, diffracted mate it was possible to index the contours as shown. In this case, the electron beam was approximately along a [100] zone axis of the crystal. A conventional electron diffraction pattern of this specimen is shown in Fig. 16.

Closer examination of the extinction contours of Fig. 14 shows that they exhibit sharp changes in direction and discontinuities hardly explainable on the basis of Eq. (33). That is, simple bending or thickness change alone does not satisfactorily account for the fine details of the contours. Highly localized distortions or orientation changes are apparently required in order to explain the details of the contours. Little more can be said at this time concerning local distortions. A sequence of shadow microscope images taken at small angular intervals should make it possible to map out the orientation of the crystal in the vicinity of a contour discontinuity.

Quite a number of electron micrographs of annealed pure aluminum sections have been obtained over the last year. A great variety of contour patterns have been observed indicating a wide variation in thickness and degree of bending in the sections. Well-annealed sections sometimes show weak Kikuchi lines by electron diffraction but not as consistently or as strong as those obtained by reflection from single crystals. This is in line with the occurrence of the extinction contours which are a direct indication of distortion since an undistorted crystal slab should be free of the contours. It is suggested that a study of the fine details of the contours in sections plastically deformed under controlled conditions may yield important information concerning dislocations.

FIG. 19. Enlargement of a region of Fig. 18 showing the granular nature of the domains in cold-worked aluminum. The size of the granules varies from about 200A to about 800A.

The preparation procedure will have to be refined, however, to reduce the amount of bending and distortion. The case of large amounts of deformation will be taken up next.

D. Structure of Cold-Worked Aluminum

The results obtained with heavily cold-worked thin sections are of considerable interest and in sharp contrast to those discussed under C. These specimens were prepared by first annealing the rolled sheet, punching the $\frac{1}{8}$-in. disks and then cold-working the central area of the disk by pounding on a small anvil.

Figure 17(a) illustrates the general type of image obtained from cold-worked aluminum and Fig. 17(b) the conventional electron diffraction pattern. Figure 17(a) shows what appear to be grains or domains of the order of 1–2μ in diameter. Stereomicrographs of this type of structure show, in general, that the domains differ in orientation and that they diffract pretty much as a whole. Figure 18 illustrates this effect.

There are several experimental conclusions pertinent to the domain structure.

(a) The domain structure of Fig. 17(a) is typical of cold-worked aluminum.
(b) The size of the domains varies from 1–3μ and is not related to the grain size revealed by etching.
(c) The domains are not revealed by etching since there is no preferential etch attack that can be observed by an anodic oxide replica.
(d) Extinction contours due to bending have rarely been observed in the domain structure. In one instance, however, a set of fringes at an edge were observed in a domain. The contours ended abruptly at the domain boundaries.
(e) Annealing of cold-worked sheet results in a gradual removal of the domain structure as annealing time increases. The domains do not appear to grow in size throughout the specimen but, rather, disappear in patches to give metal that shows the extinction contours of Fig. 14. In an intermediate state of anneal, the specimen contains regions such as Fig. 17 along with those typified by Figs. 13 and 14. The extent to which recrystallization and grain growth are involved in the disappearance of domains is not clear from the experiments performed to date. Larger usable areas of the section will be necessary in order to study recrystallization.

In addition to the relatively large domains evident in Fig. 18, a fine granular structure can be seen. It was thought at first that the granular appearance was due to oxide on the surface of the section. Since it was always observed in the cold-worked specimen but not in the annealed ones, and since the details change greatly with orientation of crystal and electron beam, the oxide explanation is unlikely. This was born out in the case of an annealed section bearing a granular oxide film in which the granular detail was easily recognizable in both members of a stereo pair. The alternative explanations are either fine scale thickness variations or very small disoriented blocks. An enlargement of a region of Fig. 18 is shown in Fig. 19 where the granular nature is quite evident. The granularity is practically absent in sections that have been partially annealed or allowed to stand some time at room temperature. This observation, along with the absence of the extinction contours, strongly suggests that the cold-worked metal initially consists of disoriented regions about 200A in size. These regions

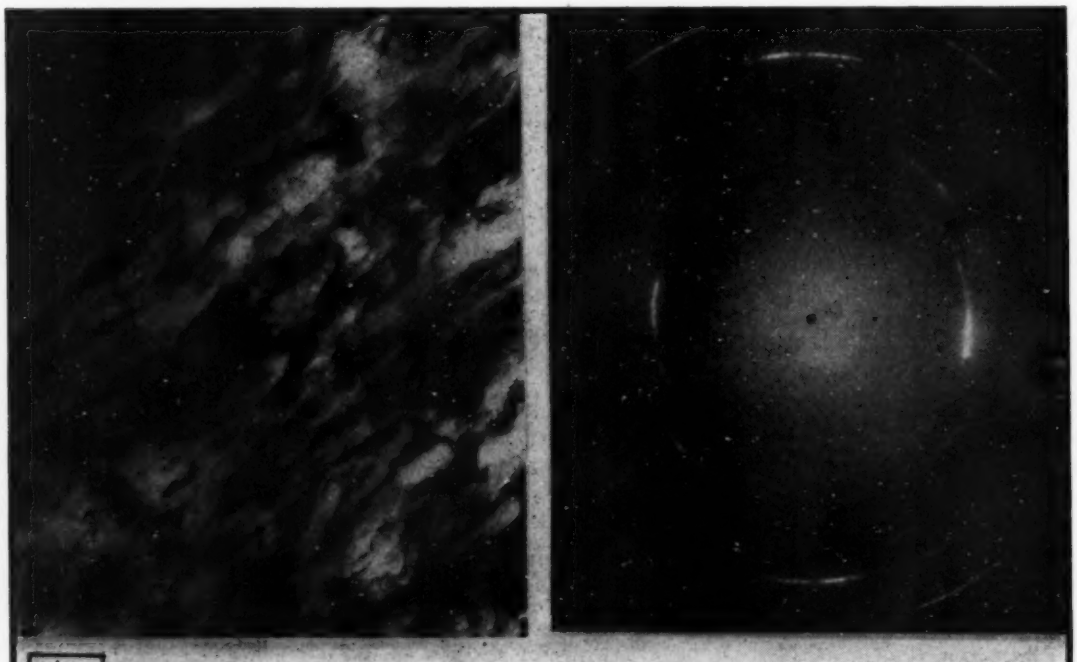


FIG. 20. Electron micrograph and diffraction pattern from thin section of as-rolled Al-4 percent Cu alloy.

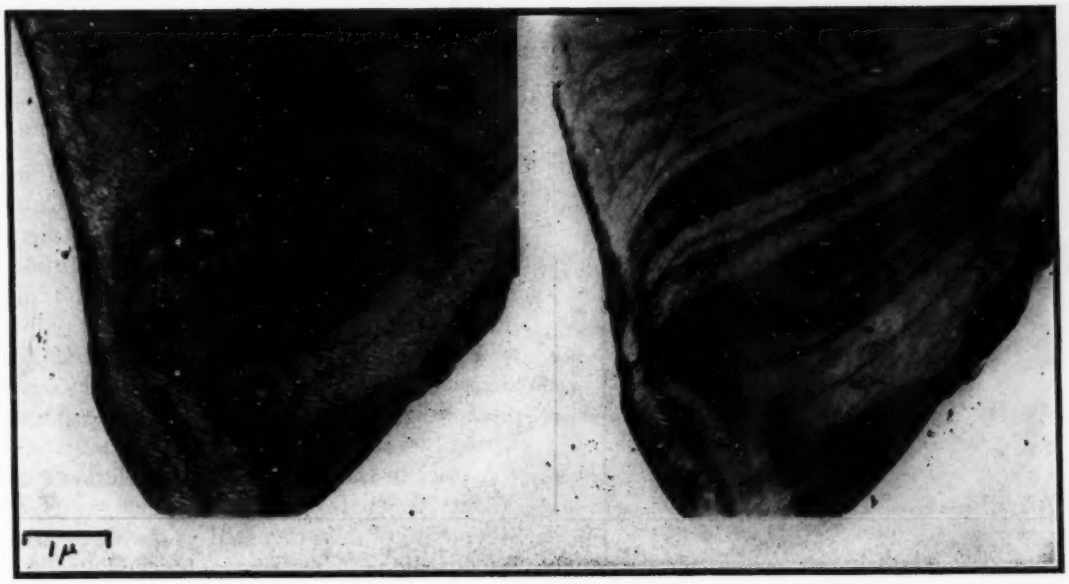


FIG. 21. Stereomicrograph of as-quenched Al-4 percent Cu section showing extinction contours and small-scale laminar structure.

scatter incoherently and hence the intensities rather than the amplitudes are additive.

The stepwise slip mechanism (reference 5) offers a natural explanation for the small blocks or regions resulting from cold-work. During working several slip systems become active to produce laminar slip bands. The intersection of the various slip bands would then produce small, disoriented prisms about 200A in size. Frank[e] has advanced a theory of the laminar slip bands based on the dynamical behavior of dislocations. This theory accounts for the relative rotation of adjacent lamellae by means of screw dislocations and for the thickness of the lamellae (about 100 atoms) by inhomogeneous strains in the immediate neighborhood of the dislocation. On this basis, then, the disoriented blocks observed in the cold-worked metal are the result of a large system of entangled dislocations with the blocks themselves inhomogeneously strained. Such strains would invalidate the assumptions of the dynamical theory and consequently dynamical effects would not be expected. The kinematical theory applied to a reciprocal lattice whose points are spread into regions should be adequate in this case.

The free energy of the cold-worked metal is considerably higher than that of the annealed metal because of inhomogeneous strains and the energy of the dislocations defining the blocks. Such a state is relatively unstable and might be expected to tend toward a less unstable structure. One way in which this change could come about would be by a reorienting of the small blocks and the resultant relief of heterogeneous strains. Reorientation would begin at discrete nuclei scattered through the volume and then proceed from these centers. Since all nuclei would not initiate precisely the same orientation, the final result would be the larger, disoriented domains of Fig. 17. The reorientation of the blocks within the domain in general would not be perfect unless a long period of time were allowed for the process to take place. The number of large domains (1–2μ) would then be the number of reorientation nuclei.

It is now suggested that the process of reorientation of the small blocks, and partial relief of internal stresses resulting therefrom, is the structural change bringing about self-recovery in cold-worked metal. Recrystallization with the elimination of the domains would bring

FIG. 22. The same section as Fig. 21 after precipitation heat treatment in vacuum at 250°C. The precipitate particles appear as dark streaks or patches.

[e] See reference 5, p. 46.

FIG. 23. The same section as Fig. 22 showing partial solution of precipitate particles by vacuum heat treatment at a higher temperature.

about complete recovery. This entire process can be viewed as a result of the dissolution and movement of dislocations.[18] The change in mechanical properties accompanying reorientation and domain formation can be qualitatively followed using Bragg's theory.[d] The model for cold-worked metal suggested here should account for x-ray line breadth observations (particle size and strains) and for low angle x-ray results.

E. Aluminum-Copper Alloy Precipitation

A final example of the application of thin sections concerns precipitation in an Al-4 percent Cu alloy. The alloy was made up using the same base metal as was employed in the preceding work. The starting point in the work to be described was the 0.005-in. rolled sheet. A thin section of the as-rolled alloy was prepared from which an electron micrograph and diffraction pattern are shown in Fig. 20. The preferred orientation is obvious in the diffraction pattern and the fragmentation caused by rolling apparent in the micrograph.

A piece of the rolled sheet was solution heat-treated (about 500°C) for a time and then water quenched. Sections were then prepared as described in A. A particular region of a section was chosen in which to follow the effect of vacuum heat treatment on the structure. Figure 21 shows stereo electron micrographs of the chosen area in the as-quenched condition.[e] The extinction contours for a bent crystal are in evidence in Fig. 21. In addition, it will be noted that one of the images exhibits a laminar appearance. This was thought to be a spurious effect at first until it was found that it was visible only for a certain orientation of electron beam and section. It was found in all sections of the as-

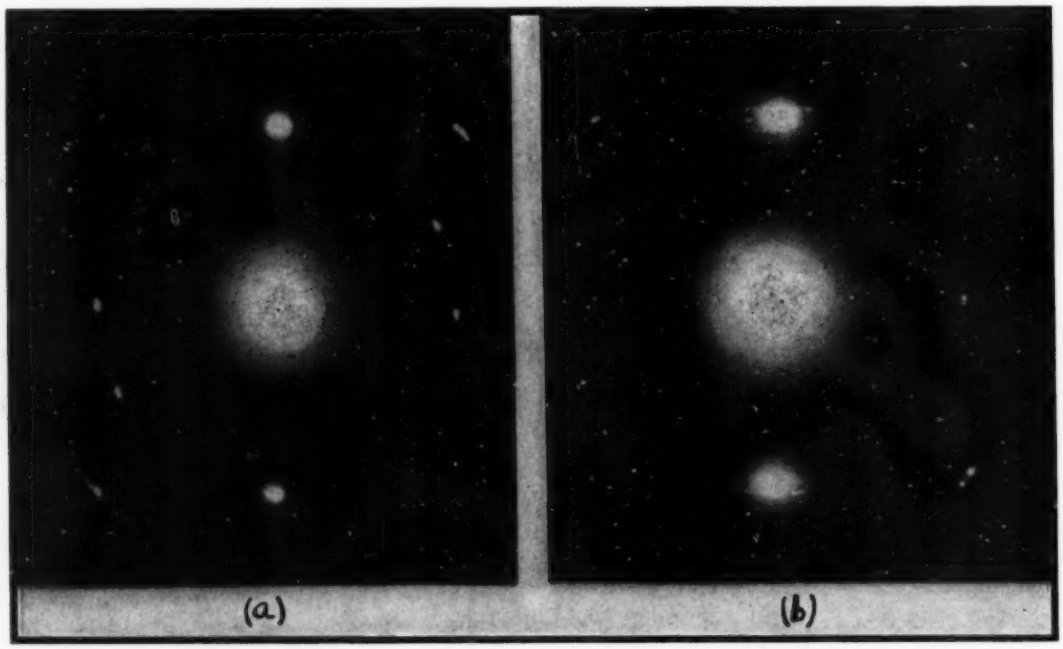


FIG. 24. Electron diffraction patterns from Al-4 percent Cu section. (a) as-quenched: corresponds to Fig. 21. (b) After precipitation of $CuAl_2$; corresponds to Fig. 22.

[18] W. G. Burger, Konink. Nederland. Akad. Weten. L, Nos. 5, 6, 7, and 8 (1947).
[d] W. L. Bragg, Trans. East Coast I. E. S. 62, 25 (1945).
[e] There is undoubtedly some local heating by electron bombardment during examination in the microscope.

quenched alloy if the orientation was correct but was not found in cold-worked or slow-cooled specimens.

Following Fig. 21, the section was vacuum heat-treated at 250°C for about ½ hour to bring about precipitation. The appearance of the precipitate is illustrated in Fig. 22 which is exactly the same area as Fig. 21. The precipitated particles are seen to be elongated in shape and are apparently platelets oriented quite accurately parallel. Near one edge of the section, which has curled, the plate shape of the precipitate is evident.

The section was again vacuum heat-treated at a considerably higher temperature where the precipitate partially dissolved as seen in Fig. 23. That the precipitate is $CuAl_2$ was verified by electron diffraction. Figure 24 shows the diffraction patterns of the as-quenched and of the precipitated section.

The striations mentioned in conjunction with the as-quenched alloy were preserved through the subsequent heat treatments although it is difficult to ascertain whether they are identically the same. Their effect on the extinction contours and the uniqueness of their appearance in quenched material is good evidence that they are "real." Whether or not they are associated with a rhythmic segregation of Cu atoms or Preston-Guinier zones cannot be said on the basis of the limited work to date.

SUMMARY AND CONCLUSIONS

(1) The dynamical theory of electron diffraction is treated using the approximation of nearly free electrons and Brillouin zones employed in the band theory of solids. The intensities of diffraction are calculated using Bloch one-electron wave functions in the crystal for the case of two strong crystal waves; a transmitted and a diffracted wave.

Although nothing essentially new results from this treatment of the theory, the description of diffraction phenomena using Brillouin zones appears advantageous. This is particularly true in the case of more than two crystal waves wherein the problem becomes one of calculating the energies near a Brillouin zone corner where perturbation of the energy levels gives rise to multiple energy gaps. As a result, a forbidden reflection, such as the (222) in the case of silicon, may arise.

(2) The expression for the diffracted intensity predicts a periodic variation with the thickness of the crystal and with deviation from the Bragg angle. In electron images formed under the conditions of Fig. 1, these interference effects are termed "extinction contours." They occur with a crystal of variable thickness or in bent crystals as demonstrated in the images of thin aluminum sections.

(3) An electrolytic method of preparing thin sections of metals is described and its application to aluminum and an aluminum-copper alloy is demonstrated. The technique is only in a crude state of development but appears promising.

(4) Extinction contours in the images of annealed sections indicate that local distortions are present which may be the result of the passage of dislocations. Electron shadow microscope work offers the chief hope of indexing and analyzing the electron images.

(5) New information concerning the structure of cold-worked aluminum is obtained by means of the thin sections.

The results are interpreted in terms of fragmentation of a metal crystal by cold-work to produce blocks about 200A in size which are inhomogenously strained and disoriented. By means of a nucleation and growth process reoriented domains 1–3μ in size are produced. This process is suggested as the structural change responsible for the recovery of cold-worked metal.

(6) The precipitation of $CuAl_2$ in a thin section of an Al-4 percent Cu alloy is demonstrated by electron micrographs and diffraction patterns. In addition to the $CuAl_2$ platelets formed by precipitation, a fine, lamellar structure has been found in the alloy after quenching and persists through the subsequent heat treatments. The origin of this structure is unknown at present. It may be connected with copper-rich regions in the crystal or Preston-Guinier zones.

(7) Further work in improving the thinning technique to obtain larger usable areas of the section is required. When this is done, work dealing with grain boundaries, slip, cold-work domains, and recrystallization can be more effectively carried on to obtain new information. The necessary extension of the work to other metals and alloys is obvious.

The author wishes to express his thanks and appreciation to Dr. C. Herring for aid in developing the theory, to Dr. W. Shockley for valuable suggestions and to Dr. James Hillier of RCA Laboratories, Princeton, New Jersey, for the shadow microscope experiments.

APPENDIX I

Consider an infinite, perfect crystal some lattice point of which is chosen as origin. A vector to an arbitrary point in the unit cell at the origin can be written

$$r=\xi_1 a_1+\xi_2 a_2+\xi_3 a_3, \tag{Ia}$$

where the ξ's are fractions ranging from zero to unity and a_1, a_2, a_3 are the basis vectors. The vector g to a point of the reciprocal lattice is

$$g=g_1 b_1+g_2 b_2+g_3 b_3, \tag{Ib}$$

where g_1, g_2, and g_3 are integers and the b's are determined by the condition $(b_j\cdot a_i)=\delta_{ij}$; $\delta_{ij}=0$ if $i\neq j$ and $\delta_{ij}=1$ if $i=j$. Consequently, $|b_j|=1/|a_j|$ and $|g|=1/d$ where d is the interplanar spacing.

The Fourier series representing the potential in the crystal is

$$V(x,y,z)=-\Sigma_{g_1}\Sigma_{g_2}\Sigma_{g_3} v_{g_1g_2g_3}e^{2\pi i(g\cdot r)},$$

where $v_{g_1g_2g_3}$ is the Fourier coefficient of potential. Removing the term $g_1=g_2=g_3=0$ leads to Eq. (5).

The coefficients v_g can be calculated from the net charge density distribution in the crystal[f] using Poisson's equation to give

$$v_g=(e/\pi\Omega|g|^2)\,\Sigma_j\,(Z_j-f_j)e^{2\pi i(g_j\cdot r_j)}\ \text{e.s.u.} \tag{Ic}$$

[f] See reference (8), Appendix 7.

In IC, e is the electron charge, Ω the volume of the unit cell, Z_j the atomic number of atom species j, and f_j the atom form factor. Here $r_j=u_j+v_j+w_j$ where (u_j, v_j, w_j) are the coordinates of atom j. In terms of the Miller indices (hkl); $(g_j\cdot r_j)=hu_j+kv_j+lw_j$. If v_g is expressed in volts, it will be written V_{hkl} so that $V_{hkl}=300v_g$ (as given by IC).

Structures not possessing a center of symmetry result in Fourier coefficients of potential that are complex which requires that $v_g^*=v_{-g}$ if $V(x, y, z)$ is to be real.

APPENDIX II

The secular Equation for (9) with $V_\lambda=0=V_{\lambda-g}$ yields the roots

$$E=E^0\pm\left\{|V_s|^2+(|V_g|^2/2)\pm\left[|V_s|^2\left(|V_g|^2+\frac{|V_s|^2}{4}\right)\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}},$$

for the special case $|K_0|^2=|K_g|^2=|K_s|^2=|K_\lambda|^2$. It is unlikely that this condition will be fulfilled in an actual lattice. However, if $(|K_0|^2-|K_g|^2)$ is quite large at the corner formed by Brillouin zones s and λ, then C_g can be neglected and the secular equation reduced to a cubic. The cubic yields three roots at the zone corner and two distinct energy gaps. A calculation for the germanium (331)|(222) zone corner indicates that an intense (222) reflection will result from the perturbation of the energy levels at the zone corner. This perturbation arises through the Fourier coefficients V_g and V_s and is operative over the entire (222) zone boundary although the effect is most pronounced near the corners. There is thus, a variable, multiple energy gap at the (222) zone boundary in the diamond type lattice.

APPENDIX III

Boundary Conditions

Let the incident surface of the crystal slab be the $x-y$ plane and K_0 the incident vacuum wave vector. If ψ_0 is the wave function for the incident electron in vacuum, then at $z=0$[8] the conditions insuring a smooth fit between the vacuum and crystal waves are

$$(\psi_0)_{z=0}=(\psi_c)_{z=0};\ \left(\frac{\partial\psi_0}{\partial z}\right)_{z=0}=\left(\frac{\partial\psi_c}{\partial z}\right)_{z=0}. \tag{IIIa}$$

In the case of plane waves, this boundary condition requires that the tangential components of all wave vectors either be equal or differ only by the tangential component of the reciprocal lattice vector.

The vacuum wave on the incident side of the crystal will in general consist of an incident and reflected wave of the form

$$\psi_0=Ae^{i(K_0\cdot r)}+Be^{i(\bar{K}_0\cdot r)},$$

where $\bar{K}_0$ is the reflected wave vector. In this case the reflected wave is not a diffracted wave but one reflected by the potential barrier at the surface. The reflection coefficient for such a barrier will be approximately

$$\left|\frac{B}{A}\right|^2=\left(\frac{(E+V_0)^{\frac{1}{2}}-(E)^{\frac{1}{2}}}{(E+V_0)^{\frac{1}{2}}+(E)^{\frac{1}{2}}}\right)^2, \tag{IIIb}$$

which is negligible for fast electrons. Consequently, only the incident vacuum wave $\psi_0=Ae^{i(K_0\cdot r)}$ need be considered for the case of nearly normal incidence. Remembering that K and K' correspond to incident crystal waves and K_g and K_g' to diffracted crystal waves, Eqs. (18) and (19) yield the relations

$$\begin{aligned}\frac{b}{C_{K0}((1+\alpha_g^2)\Omega)^{\frac{1}{2}}}+\frac{a}{C_{K'0}((1+\alpha_g'^2)\Omega)^{\frac{1}{2}}}&=A,\\ \frac{\alpha_g a}{C_{K0}((1+\alpha_g^2)\Omega)^{\frac{1}{2}}}+\frac{\alpha_g' b}{C_{K'0}((1+\alpha_g'^2)\Omega)^{\frac{1}{2}}}&=0.\end{aligned} \tag{IIIc}$$

The first relation equates the transmitted crystal and incident vacuum amplitudes and the second equates diffracted amplitudes to zero at $z=0$. The exponentials do not appear since the tangential components of all wave vectors are equal at $z=0$. Solving these equations for a and b gives

$$\begin{aligned}a&=\frac{\alpha_g'((1+\alpha_g^2)\Omega)^{\frac{1}{2}}}{\alpha_g'-\alpha_g}C_{K0}A,\\ b&=-\frac{\alpha_g((1+\alpha_g'^2)\Omega)^{\frac{1}{2}}}{\alpha_g'-\alpha_g}C_{K'0}A.\end{aligned} \tag{IIId}$$

The wave function for the diffracted wave, Eq. (22), follows immediately.

[8] It is assumed that the periodic potential vanishes abruptly at $z=0$. Von Laue, Phys. Rev. 37, 53 (1931) has shown that this is a good approximation for fast electrons.

Letters to the Editor

Erratum: Determination of Relative Phases of Fourier Coefficients from X-Ray Diffraction Data

[J. App. Phys. 20, 295 (1949)]

DAN MCLACHLAN, JR.
Department of Metallurgy, University of Utah, Salt Lake City, Utah

SINCE the publication of a recent Letter to the Editor, it has been courteously pointed out by Dr. David Harker and co-workers that a lack of generality exists in Eq. (2) of that letter.

It is the feeling of the author that this error should be made known to possible users of the equation. The correction of the equation, however, is so involved that it is perhaps wise to postpone the full discussion of it until a more lengthy paper can be prepared.

Increase in Q-Value and Reduction of Aging of Quartz Crystal Blanks

A. C. PRICHARD, M. A. A. DRUESNE, AND D. G. MCCAA
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
July 22, 1949

THE importance of quartz crystal units for frequency control derives from the very high Q (ratio of reactance to resistance) and the low frequency-temperature coefficient of quartz, and the very steep slope of the reactance curve of the crystal blank. The Q of quartz is higher than that of any other frequency controlling device used to date.

The Q and frequency of a given quartz crystal blank change, however, with time; these changes are results of the so-called aging process. During World War II, crystal units were used in very large quantities in military communications equipment. Considerable difficulty was experienced during the first years of the war because the useful life of a crystal unit was greatly reduced by reason of its "aging out of frequency" or "going dead." A usable reduction in aging was later accomplished during 1944 by etching the surfaces of the crystal blanks. This only mitigated the effect, however, and its more complete elimination was highly desirable.

A method has recently been developed at the Signal Corps Engineering Laboratories, Fort Monmouth, N. J. whereby the Q-value of the quartz blank has been materially increased, and the aging has been very greatly reduced. The method consists of annealing the quartz blank by heating it almost to the inversion temperature of quartz, or 500°C, and cooling it down extremely slowly.

This simple treatment yields values of Q which are at least double those of untreated quartz blanks, and variations in frequency and Q are minute. These improvements in Q and aging characteristics appear to be of a permanent character.

A paper describing the results in greater detail is in preparation.

Addendum: Analysis of the Metal-Strip Delay Structure for Microwave Lenses

[J. App. Phys. 20, 257 (1949)]

S. B. COHN
Radio Engineering, Sperry Gyroscope Company, Division of the Sperry Corporation, Great Neck, New York

Since the graphs in the above-titled paper are difficult to read, the numerical values are presented in Table I.

TABLE I. Index of refraction and image admittance values for the metal-strip delay lens

l/b	b/λ	$\delta=0.3$ η	$\delta=0.3$ y_I	$\delta=0.4$ η	$\delta=0.4$ y_I	$\delta=0.5$ η	$\delta=0.5$ y_I
$3/4\pi$	0	1.536	1.536	1.361	1.361	1.237	1.237
	0.3	1.646	1.388	1.432	1.270	1.280	1.182
	0.4	1.770	1.233	1.505	1.175	1.322	1.125
	0.5	2.055	0.9518	1.657	1.017	1.405	1.036
$1/\pi$	0	1.490	1.490	1.337	1.337	1.225	1.225
	0.3	1.583	1.325	1.399	1.236	1.263	1.164
	0.4	1.697	1.152	1.465	1.133	1.301	1.103
	0.5	1.947	0.8115	1.603	0.9454	1.370	0.9971
$3/2\pi$	0	1.400	1.400	1.279	1.279	1.188	1.188
	0.3	1.475	1.187	1.325	1.148	1.216	1.108
	0.4	1.567	0.9402	1.376	1.003	1.243	1.028
	0.5	1.924	0.2195	1.488	0.7109	1.299	0.8669
$2/\pi$	0	1.328	1.328	1.229	1.229	1.154	1.154
	0.3	1.394	1.047	1.267	1.056	1.176	1.048
	0.4	1.504	0.6490	1.316	0.8385	1.200	0.9229
	0.5		imaginary values			1.253	0.6155

These values were calculated from Eqs. (16), (18), (24), and (25) of the above-named paper, using the parameters defined in it.

Surface Polarization and Wettability of Inorganic Substances by Water*

D. P. ENRIGHT, L. R. SONDERS, AND W. A. WEYL
The Pennsylvania State College, State College, Pennsylvania
June 22, 1949

IN a recent publication M. K. Roman, E. C. Marboe, and W. A. Weyl[1] describe the influence of heavy metal ions on the hygroscopicity of a glass surface. The surface forces of a glass can be greatly decreased by bringing into the surface layer cations with incomplete outer electronic shells. Ions of this type, e.g., Pb^{2+} or Tl^{+}, are strongly polarized in the asymmetrical force field of the glass-air interface, i.e., they extend strong forces toward the interior of the glass and only weak forces toward the gas phase. A theoretical explanation of the surface properties of glasses as affected by heavy metal ions has been given by W. A. Weyl[2] and a method for estimating surface forces between solids and liquids has been developed by E. C. Marboe and W. A. Weyl.[3]

Further experiments with inorganic materials and their wettability by water showed that the polarization of the surface ions is one of the major causes of the "receding contact angle." N. K. Adam[4] writes: "The cause of the hysteresis of contact angles is still obscure." A recent symposium held by the Faraday Society devoted considerable discussion to this subject without making much progress. It was pointed out[5] that contamination, surface roughness, and orientation of polar molecules may be responsible for the fact that many surfaces give different contact angles with water depending on whether or not the surface had been previously wetted. There can be no doubt that the structure of the surface, the adsorbed gas film, contamination of the surfaces by greasy material, etc., affect the contact angle. Obviously, the "turning around of molecules" is restricted to those solids which consist of polar molecules (organic compounds) or to these contaminated by organic molecules. For many inorganic compounds such as $CrCl_3$ and CdI_2 the term molecule cannot be applied, but the phenomenon is the same. The polarizability of an ion may be described as its ability to adjust its force field to its environment. Whereas the force fields of small noble-gas-type ions can be conveniently derived from their charges and radii, this cannot be done for large ions (I^-, S^{2-}) or ions with incomplete outer electron configuration (Pb^{2+}, Hg^{+}, Tl^{+}). The electron configuration of these ions is strongly affected by the electrical fields of their surroundings. The deformation of ions is particularly strong in asymmetrical force fields as was demonstrated by K. Fajans.[6] The

surface energy of substances containing strongly polarizable ions is lowered by bringing these ions into the surface layer.[2]

The incomplete wettability of soluble inorganic substances by water is the result of the weak forces emanating from the strongly polarized surface ions. The change of their wettability in contact with liquid water is the result of the contrapolarizing influence of the liquid. It is immaterial whether or not this contrapolarization and hydration leads to solution.

The following experiments may illustrate the importance of the polarization of surface ions on the wettability.

1. Anhydrous chromic chloride or cadmium iodide forms layer structures containing the large and therefore polarizable anions in a state of strong deformation. The "greasy feel" of these materials and their good cleavage are indicative of weak force fields. Compounds of this type are not completely wet by water. In spite of its solubility, $CrCl_3$ crystals may float on water for hours. In the presence of acids, due to the strong contrapolarizing influence of the protons on the deformed Cl^- ions, the surface ions come into a more symmetrical force field; they become wettable and finally hydrated. This puzzling behavior is well known to be characteristic for many water soluble substances. That contamination is not the answer to the problem can be seen from the following experiments.

2. Mercuric iodide contains the strongly polarizable ions Hg^{2+} and I^-. When precipitated from aqueous solutions, the surface ions of the HgI_2 are partly hydrated and as a result find themselves in fairly symmetrical force fields. If the precipitate is filtered by means of a suction funnel, it behaves like other hydrophylic substances. However, the moment the precipitate has been allowed to come in contact with air, the surface polarization reduces its wettability. After the supernatant liquid has been removed by suction, the next drop of water no longer wets the mercuric iodide completely, even if the elapsed time is not more than a few seconds. As this time is too short for contamination by greasy materials, and as the mercuric iodide does not contain molecules which can "turn around," it leaves only the surface roughness as a possible explanation for this striking hysteresis of the contact angle. That the surface roughness is not the cause can be seen from the following experiment.

3. Clay minerals, especially montmorillonite, are characterized by their base exchange properties. The acid or hydrogen clay in contact with aqueous solutions of salts exchanges the H^+ for other cations such as Na^+, Ca^{2+} or Pb^{2+}. By treating montmorillonite with different salts, suspensions of various heavy-metal clays were obtained. These clay slips were sprayed on unglazed ceramic tiles and allowed to dry in a drying oven. All clay deposits had the same chance to become contaminated and had identical surface roughness. Nevertheless, the Na^+ clay was found to absorb a drop of water immediately. The Ca^{2+} clay, due to the greater polarizability of the Ca^{2+}, showed a distinctly poorer wettability but still absorbed a drop of water brought onto its surface within 4–5 seconds. Pb^{2+} clay and clay treated with other highly polarizable cations was not completely wet by water. Water brought in contact with heavy-metal ion clay showed contact angles of about 70° and was not absorbed until a considerable time (up to 15 minutes) had elapsed.

4. Further experiments on the rise of water in capillaries of a soda-lime glass which first had been treated with aqueous solutions of heavy-metal salts and then allowed to dry confirmed our explanation without contributing additional information.

These experiments, which shall be described in more detail elsewhere, lead to an explanation of the hysteresis of the contact angle. It shall not be denied that for organic substances the orientation of surface molecules may have an effect equally as strong as the polarization of surface ions in the asymmetrical force field of the crystal-gas interface.

* Sponsored by the ONR and carried out under Contract No. N6 onr 269 Task Order 8 NR 032-265.

[1] Roman, Marboe, and Weyl, J. Soc. Glass Tech. 32, 260–263 (1948).
[2] W. A. Weyl, J. Soc. Glass Tech. 32, 247–259 (1948).
[3] E. C. Marboe and W. A. Weyl, J. Soc. Glass Tech. 32, 281–290 (1948).
[4] N. K. Adam, *The Physics and Chemistry of Surfaces* (Oxford University Press, New York), third edition, p. 180.
[5] N. K. Adam, Trans. Faraday Soc. No. 3, 5–11 (1948).
[6] K. Fajans, J. Chem. Phys. 9, 373 (1941).